UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadística e Investigación Operativa

TESIS DOCTORAL

Pérdida de información a causa de la censura

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Agustín Turrero Nogues

DIRECTOR:

Julián de la Horra Navarro

Madrid, 2015

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Matemáticas

Departamento de Estadística e Investigación Operativa

PERDIDA DE INFORMACION A CAUSA DE LA CENSURA

Agustín Turrero Nogués

Madrid, 1988

Colección Tesis Doctorales. N.º 361/88

Edita e imprime la Editorial de la Universidad
Complutense de Madrid. Servicio de Reprografía
Noviciado, 3 - 28015 Madrid
Madrid, 1988
Ricoh 3700
Depósito Legal: M-26544-1988

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMATICAS

Sección de Estadística e Investigación Operativa

PERDIDA DE INFORMACION A CAUSA DE LA CENSURA

AGUSTÍN TURRERO NOGUÉS

Memoria para optar al grado de
Doctor en Ciencias Matemáticas
realizada bajo la dirección del
Dr. D. Julián de la Horra Navarro

Madrid, Noviembre de 1987

A Pilar García-Carrasco

CONTENIDO

PRESENTACION

PRESENTACION

Exponemos esta memoria en un texto principal y un apéndice.

El texto está dividido en tres capítulos, los dos primeros teóricos y el tercero de aplicación práctica de los anteriores.

La aportación original de este trabajo se desarrollará especialmente en los dos primeros capítulos.

En el capítulo I proponemos tres nuevas medidas paramétricas de información reales basadas en la matriz de información de Fisher.

En el capítulo II proponemos dos posibles alternativas para evaluar la pérdida de información que, acerca de un parámetro, se produce en modelos de supervivencia censurados.

En el último capítulo introducimos de manera intuitiva un modelo de supervivencia censurado que nos servirá para valorar y dotar de contenido práctico las definiciones y propiedades expuestas en los capítulos precedentes.

El apéndice recoge aquellos resultados no elementales que sirven de apoyo al texto principal.

Continuamos esta presentación ofreciendo un resumen introductorio sobre el contenido de esta memoria.

Resumen

El análisis de supervivencia es un concepto estadístico amplio que abarca una gran variedad de técnicas estadísticas para el análisis de variables aleatorias positivas. Generalmente, la variable aleatoria es el tiempo hasta el primer fallo de una com

ponente física (mecánica o eléctrica), o el tiempo hasta la - muerte de una unidad biológica (paciente, animal, célula).

Pensamos que la distribución de la variable pertenece a una familia paramétrica y que estamos interesados en hacer inferencias acerca del parámetro θ. Como vehículo para este proceso inferencial, disponemos de un experimento estadístico en el que se observa la variable en un grupo de unidades muestrales. La - primera dificultad que nos vamos a encontrar en la mayoría de - los estudios sobre supervivencia es la imposibilidad de dispo-- ner de los valores exactos de la variable para todas las unidades muestrales debido a lo que, de modo genérico, se llama la - censura. Por ejemplo, en estudios de fiabilidad sobre la dura-- ción de una determinada componente, no será útil esperar a la - destrucción de la misma; en supervivencia, algunos pacientes - pueden sobrevivir al final del ensayo clínico o pasar a otro - tratamiento o morir por otra causa ajena a la enfermedad en estudio, produciéndose en cualquier caso una observación incompleta para algunas unidades muestrales; diremos que dichas unida-- des están censuradas.

Teóricamente pues, disponemos de dos experimentos para obtener información acerca de θ, el experimento censurado ε_c y el corrrespondiente no censurado ε_o (si no interviniera ninguna variable de censura). Si tuviésemos la posibilidad de elección entre ambos con el objetivo inferencial descrito anteriormente, - parece evidente que la elección recaería sobre ε_o.

Sin embargo la situación más frecuente no será ésta, sino la de elegir entre varios experimentos censurados ε_{c_i} $i=1,\dots,r$ (según r variables de censura distintas). Un método natural para elegir el más apropiado sería comparar la información estadística acerca de θ, contenida en dichos experimentos y elegir el de mayor información. Ahora bien, como es conocido, no hay una única medida de la información contenida en un experimento.

En un artículo aún no publicado, Goel (1987) justifica el utilizar cualquier medida de información que verifique la propiedad de suficiencia de experimentos (si un experimento ε_1 es suficiente para otro ε_2 según la definición de Blackwell (1951; 1953), entonces ε_1 contiene más información acerca del parámetro que ε_2). Con este punto de partida, nuestro objetivo será evaluar la pérdida de información que se produce al observar, o bien ε_c en lugar de ε_o, para cualquier variable de censura, o bien ε_{c_1} en lugar de ε_{c_2}, donde la variable de censura de ε_{c_2} es estocásticamente mayor que la de ε_{c_1}.

Este problema no admite complicación si la medida de información es real, pero pensemos en la situación en que el parámetro de interés θ es k-variante. Desde un punto de vista clásico, la única medida de información paramétrica adecuada es la matriz de Fisher que, a todas luces, es inoperante para el objetivo fijado.

Proponemos dos vías para abordar este problema, que desarrollamos en el capítulo II: la primera consta a su vez de dos etapas: a) definimos dos medidas matriciales de la pérdida de información que llamamos matriz de pérdidas y matriz de eficiencias que justificamos de manera formal por las propiedades de sus autovalores y de manera intuitiva por el claro significado de dichos autovalores al diagonalizar la matriz de Fisher mediante una transformación biyectiva del parámetro de interés; b) defini mos varias medidas reales de la pérdida de información debida a la censura, mediante distintas funciones reales de las matrices anteriores y en particular de sus autovalores, heredando de esta forma las buenas propiedades de éstos, destacando la propiedad de invariancia bajo transformaciones biyectivas del parámetro, para todas las medidas. Todo este proceso va a requerir que la matriz de Fisher correspondiente al experimento no censurado sea no singular.

Una segunda forma de abordar nuestro objetivo sería definir primeramente medidas reales de información basadas en la matriz de Fisher y a partir de aquéllas calcular la pérdida relativa de información y la eficiencia relativa del experimento ε_c respecto del ε_o de una manera "natural", para cada una de dichas medidas, teniendo en cuenta para ello el concepto de eficiencia relativa que utiliza Brooks (1982) en un contexto bayesiano. Esta segunda vía no requiere ninguna hipótesis restrictiva, salvo la de encontrar "las medidas reales basadas en la matriz de Fisher". Dedicamos el capítulo I a tal fin:

Proponemos tres funciones reales, monótonas crecientes y simétricas de los autovalores de la matriz de Fisher como medidas de información paramétricas reales, que justificamos intuitivamente mediante el análisis de componentes principales de dicha matriz y de una manera formal a la vista de las propiedades específicas que como medidas de información poseen.

El proceso que abarca los capítulos I y II puede sintetizarse en el siguiente esquema:

$$
\begin{tikzcd}[column sep=small]
 & \text{Pérdida Matricial de Información} \arrow[dr] & \\
\text{Matriz de Información de Fisher} \arrow[ur] \arrow[dr] & & \text{Medida Real de Pérdida de Información} \\
 & \text{Medida Real de Información} \arrow[ur] &
\end{tikzcd}
$$

Dedicamos, por último, el tercer capítulo a la aplicación práctica de los conceptos teóricos propuestos en los dos capítulos precedentes.

Introducimos de una manera intuitiva un modelo de supervivencia censurado cuya formalización induce una familia de experimentos ε_c en los que observamos una variable aleatoria de distribución multinomial cuya dimensión depende de la distribución prefijada de la variable de censura y cuya ley de probabilidad depende de un parámetro k-variante θ para toda variable de censura.

Calculamos las medidas paramétricas de información adecuadas: la matriz de Fisher y las medidas reales basadas en ella que proponíamos en el capítulo I. Evaluamos la pérdida de información debida a la censura utilizando los dos métodos propuestos en el capítulo II, comparando y discutiendo los resultados obtenidos.

Quiero finalizar esta presentación con un recuerdo emocionado para la profesora Dña. Pilar García-Carrasco Aponte que dirigió este trabajo desde su inicio y que, desgraciadamente, no

pudo ver acabado. Por su ayuda y amistad mi agradecimiento más sincero. Quiero también expresar mi gratitud al profesor D. Julián de la Horra Navarro por su dirección y apoyo desinteresados en los momentos más difíciles, así como al profesor D. Eusebio Gómez Sánchez-Manzano por las ideas y consejos recibidos.

Madrid, noviembre de 1987

CAPITULO I

NUEVAS MEDIDAS DE INFORMACION BASADAS EN LA MATRIZ DE FISHER

CAPITULO I

NUEVAS MEDIDAS DE INFORMACION BASADAS EN LA MATRIZ DE FISHER

1.0. SUMARIO

La noción de información estadística, acerca de un parámetro, contenida en un experimento, ha tenido diferentes formalizaciones en el tiempo, según diferentes autores. Presentamos una clasificación de las medidas más usuales al respecto, así como una exposición de las propiedades deseables en toda medida de información (Sección 1.1).

Suponemos que el parámetro de interés es k-variante y centramos la atención en las medidas paramétricas adecuadas, que se reducen, bajo condiciones de regularidad, a la matriz de información de Fisher. Recogemos las principales propiedades de esta medida (Sección 1.2).

Si bien la matriz de Fisher se comporta como una buena medida de información en lo que respecta a sus propiedades, adolece de una interpretación intuitiva y resulta inoperante a la hora de comparar experimentos (Apartado 1.3.1).

Una manera natural de solventar estas deficiencias sería definir una función real de dicha matriz que tuviera una interpretación intuitiva como medida de información paramétrica y como tal,buenas propiedades.

Proponemos tres funciones reales, monótonas crecientes y simétricas de los autovalores de la matriz de Fisher como medidas de información paramétricas unidimensionales, cuya justifi cación intuitiva radica en el análisis de componentes principales de dicha matriz (Apartado 1.3.2). Estas medidas, que denotamos por S_X,D_X y M_X son funcionales conocidos del álgebra matricial.

Estudiamos las propiedades de estas tres medidas, incluyendo las demostraciones de las mismas (Apartado 1.3.3). A la vista de estas propiedades concluimos el buen comportamiento de S_X, D_X y M_X como medidas de información, si bien D_X presenta algunas lagunas cuando la matriz de Fisher es singular. En este caso sugerimos utilizar una cuarta medida de información D_X^* basada en D_X que elimina aquellas deficiencias.

Posteriormente estudiamos los efectos que una transformación biyectiva del parámetro ocasiona en dichas medidas (Apartado 1.3.4), que se concretan en distintas relaciones (para S_X, D_X y M_X) entre la medida relativa al parámetro original y la correspondiente al parámetro transformado y destacando como resultado más fuerte la invariancia de todas ellas bajo transformaciones lineales ortogonales del parámetro.

Para concluir, observamos el comportamiento de las tres medidas de información respecto a dos criterios de comparación de experimentos más débiles que el de Blackwell (Apartado 1.3.5).

1.1. MEDIDAS DE INFORMACION

Denotamos por $\varepsilon_X=\{(X,\Omega_X,A);P_\theta:\theta\in\Theta\}$ al experimento estadístico que consiste en la observación de una variable aleatoria X definida en el espacio medible (Ω_X,A) cuya distribución de probabilidad P_θ depende de un parámetro desconocido θ perteneciente a un espacio paramétrico Θ. Suponemos que la familia de medidas de probabilidad $\{P_\theta:\theta\in\Theta\}$ está dominada por una medida finita o σ-finita μ. Sea $f(x,\theta)=\dfrac{dP_\theta}{d\mu}$ la densidad correspondiente.

El espacio paramétrico Θ será un subconjunto abierto de la recta real o un subconjunto abierto del espacio euclídeo k-dimensional $\mathbb{R}^k$.

Para medir la información que, acerca de un parámetro desconocido θ, proporciona un experimento ε_X se han propuesto en la literatura diferentes medidas, cada una de las cuales posee una cierta axiomática y/o determinadas propiedades. Podemos, sin embargo, agruparlas en tres categorías, las dos primeras en el marco de la estadística clásica, y la tercera en un contexto bayesiano.

a) Medidas paramétricas de información

a1) Directas

Miden la cantidad de información que proporcionan los datos, acerca de un parámetro desconocido θ, y son funciones de θ. Las principales son las medidas de Fisher (1925), Mathai (1967), Vajda (1973) y Boekee (1977) si θ es univariante; y la matriz de información de Fisher, si θ es k-variante.

a2) Indirectas (a partir de medidas no paramétricas).

Las medidas no paramétricas de información se refieren a cualesquiera familias de medidas de probabilidad $P_i \ll \mu$ con $f_i = \frac{dP_i}{d\mu}$. Miden la cantidad de información proporcionada por los datos, para discriminar en favor de una distribución f_1, en contra de otra f_2; o de otra forma, miden la divergencia o afinidad entre f_1 y f_2. Las principales son las medidas de Kullback-Leibler (1951), Rényi (1961), Csiszár (1963), Vadja (1973), Bhattacharyya generalizada (Papaioannou y Kempthorne, 1971), Kagan (1963) y Matusita (1967), pudiendo obtenerse las dos últimas a través de la de Csiszár, eligiendo convenientemente la función convexa $\emptyset$ que aparece en la definición de ésta.

Si bien este tipo de medidas no evaluan la información acerca del parámetro de interés, pues se aplican a familias no paramétricas , existen métodos que permiten construir medidas paramétricas de información, a partir de las no paramétricas anteriores (Ferentinos y Papaioannou, 1981). Así, si θ es univariante dichos autores definen las medidas de información paramétricas de Rényi y Csiszár (modificadas) y obtienen las restantes a partir de la de Csiszár, con elecciones convenientes de la función convexa $\emptyset$ de la definición de esta última.

Si θ es k-variante, definen las matrices paramétricas de Rényi y Kagan y por el mismo método se podrían definir las restantes.

b) Medidas de información en contexto bayesiano

Miden la cantidad de información, acerca de θ, que puede esperarse de los resultados de un experimento, cuando existe conocimiento a priori sobre θ. En general, son valores numéricos y por tanto no dependen de θ. Se definen, o bien como la diferencia esperada entre las incertidumbres de las distribuciones a priori y a posteriori (DeGroot, 1962), o como diferencia de la utilidad esperada de la acción bayes a posteriori y la utilidad de la acción bayes a priori (Raiffa y Schlaifer, 1961; Bernardo, 1979; DeGroot, 1984), o como "una divergencia" entre las distribuciones a posteriori y a priori.

Entre las más conocidas están la medida de información de Shannon (Shannon, 1948; Lindley, 1956) y la medida de Mallows (1959). Esta última depende del parámetro θ.

Asímismo, como muestran Ferentinos y Papaioannou (1982), todas las medidas paramétricas (directas e indirectas) de información pueden usarse en un contexto bayesiano, tomando $f(x,\theta)=P(x|\theta)$.

La importancia de una medida de información radica en sus propiedades. Si X es la variable aleatoria observada, T(X) una transformación medible de X (estadístico) e I_X una medida de información de cualquier categoría acerca de $\theta \in \Theta$, basada en X, las principales propiedades que debería poseer la medida I_X se resumen en las siguientes:

(1) No Negatividad

$$I_X \geqslant 0$$

(2) Aditividad

$$I_{X,Y} = I_X + I_{Y|X} \qquad I_{X,Y} = I_X + I_Y$$

(si X e Y son independientes)

$$I_{X,Y} \leq I_X + I_Y$$

(subaditividad)

(3) Desigualdad Condicional

$$I_Y \geq I_{Y|X}$$

(4) Información Máxima

$$I_X \geq I_{T(X)}$$

(5) Invariancia bajo Transformaciones Suficientes

$$I_X = I_{T(X)} \iff T_{(X)} \text{ es un estadístico suficiente}$$

(6) Convexidad

Sea $\alpha \geq 0$ y F familia de densidades de probabilidad con el mismo espacio paramétrico Θ

$$I_X(\alpha f_1 + (1-\alpha) f_2) \leq \alpha I_X(f_1) + (1-\alpha) I_X(f_2) \qquad f_1, f_2 \in F$$

entendiendo que $I_X(f)$ indica la información acerca de θ, proporcionada por la variable X cuya densidad es f.

(7) Suficiencia de Experimentos

Si el experimento ε_X es suficiente para el experimento ε_Y, y denotamos $\varepsilon_X \gtrsim \varepsilon_Y$, según la definición de Blackwell (1951, 1953), teniendo ambos experimentos el mismo espacio paramétrico, entonces

$$I_X \geq I_Y$$

(8) Pérdida de Información debida al Agrupamiento de Observaciones.

Sea $\Omega_X = \mathbb{R}^m$. Un agrupamiento g significará una partición $-\infty < x_{i0} < x_{i1} < \ldots < x_{in_i} < \infty$ en cada uno de los ejes i=1...m

Sea $E_{i0} = (-\infty, x_{i0})$, $E_{ij} = (x_{ij-1}, x_{ij})$ $j=2\ldots n_i$, $E_{in_{i+1}} =$

$= (x_{in_i}, \infty)$

$$y \quad E_{j_1 j_2 \ldots j_m} = E_{1j_1} \times E_{2j_2} \times \ldots \times E_{mj_m} \quad j_i = 0,1,\ldots n_{i+1}$$

Si llamamos G al conjunto de todas las particiones g de $\mathbb{R}^m$ e I_g a la medida de información basada en I_X, para la variable aleatoria g, entonces:

$$\sup_{g \in G} I_g = I_X$$

Para comprender mejor el significado de esta última propiedad, pensemos que I_X es la medida de información de Fisher y $\Theta \subset \mathbb{R}$, es decir:

$$I_X = E_\theta \left[\frac{\partial}{\partial \theta} \log f(X,\theta)\right]^2$$

Entonces:

$$I_g = \sum_{j_1 j_2 \ldots j_m} \left[\frac{\partial}{\partial \theta} \log P(E_{j_1 \ldots j_m})\right]^2 P(E_{j_1 \ldots j_m})$$

con $P(E) = \int_E f d\mu$

Las propiedades anteriores se refieren tanto al caso en que θ es univariante como k-variante. Si I_X representa una matriz paramétrica, la desigualdad $I_X \geqslant I_Y$ expresará que la matriz $I_X - I_Y$ es semidefinida positiva, y escribiremos $I_X - I_Y \geqslant 0$ o indistintamente $I_X \geqslant I_Y$. Análogamente para todas las propiedades. La propiedad (8):

$$\sup_{g \in G} I_g = I_X$$

indicará la igualdad para cada elemento, es decir:

$$\sup_{g \in G} I_g^{ij} = I_X^{ij} \qquad i,j=1,\ldots,k$$

siendo I_g^{ij} e I_X^{ij} los elementos (i,j) de las matrices I_g e I_X respectivamente.

En general, el término matriz de Información significará una medida paramétrica de información, que para cada $\theta \in \Theta \subset \mathbb{R}^k$, es una matriz k×k, simétrica real semidefinida positiva que satisface la propiedad $I_X(\theta) \geq I_{T(X)}(\theta)$ para todo estadístico T(X) y todo $\theta \in \Theta$ con la igualdad si y sólo si T(X) es suficiente para θ.

En el ámbito de las medidas paramétricas de información podemos añadir una diferencia entre las directas y las indirectas.

Las medidas paramétricas directas son aplicables a familias de distribuciones regulares, es decir a familias que satisfacen las condiciones de regularidad de la medida de Información de Fisher (Kagan, Linnik y Rao, 1973, p. 275).

Por otra parte, si θ es univariante, las medidas paramétricas indirectas se aplican a familias más amplias que incluyen las no regulares, pues sus definiciones y propiedades no requieren condiciones de regularidad. Sin embargo, si imponemos las condiciones de regularidad de la medida de información de Fisher estas medidas se convierten en funciones lineales de la medida de Fisher (Ferentinos y Papaioannou, 1981).

Si θ es k-variante, las medidas paramétricas indirectas son matrices paramétricas, que no son en general matrices de información, excepto en el caso de la matriz de Kagan y para familias regulares.

Ampliando ligeramente las condiciones de regularidad, Ferentinos y Papaioannou (1981) muestran la igualdad de las matrices de información de Kagan y Fisher. En el resto del capítulo, supondremos que θ es k-variante y estudiaremos en profundidad las medidas paramétricas de información adecuadas así como sus propiedades.

1.2. MATRIZ DE INFORMACION DE FISHER

Si θ es k-variante, como vimos anteriormente, bajo condiciones de regularidad, la matriz de información de Fisher es - la única medida paramétrica de información válida para este caso.

Sea $\theta = (\theta_1, \theta_2, \ldots, \theta_k) \in \Theta \subset \mathbb{R}^k$

Supongamos que:

A.- $f(x,\theta) > 0 \quad \forall x \in \Omega_X, \quad \forall \theta \in \Theta$

B.- $\frac{\partial}{\partial \theta_i} f(x,\theta)$ existe $\forall x \in \Omega_X, \quad \forall \theta \in \Theta \ , \ i=1,\ldots,k$

C.- Para todo $A \in \mathcal{A}$ $\quad \frac{\partial}{\partial \theta_i} \int_A f(x,\theta)\,d\mu = \int_A \frac{\partial f(x,\theta)}{\partial \theta_i}\,d\mu \quad \forall \theta \in \Theta$

$i=1,\ldots,k$

Definición

En las condiciones A, B y C anteriores, se define la matriz de información de Fisher $I_X(\theta)$, como la matriz real de orden k, cuyo elemento (i,j) es:

$$I_X^{ij}(\theta) = E_\theta \left[\frac{\partial}{\partial \theta_i} \log f(X,\theta) \frac{\partial}{\partial \theta_j} \log f(X,\theta) \right] =$$

$$= \int_{\Omega_X} \left(\frac{\partial}{\partial\theta_i} \log f \right) \left(\frac{\partial}{\partial\theta_j} \log f \right) f(x,\theta)\, d\mu =$$

$$= \mathrm{Cov}_\theta \left[\left(\frac{\partial}{\partial\theta_i} \log f \right), \left(\frac{\partial}{\partial\theta_j} \log f \right) \right]$$

$$i,j=1,\ldots,k$$

Si exigimos además:

D.- Para todo $A \in \mathcal{A}$ $\quad \frac{\partial^2}{\partial\theta_i \partial\theta_j} \int_A f(x,\theta)\, d\mu =$

$$= \int_A \frac{\partial^2}{\partial\theta_i \partial\theta_j} f(x,\theta)\, d\mu \qquad \forall\theta \in \Theta$$

$$i,j=1,\ldots,k$$

entonces, los elementos de la matriz de Fisher se pueden obtener según la expresión

$$I_X^{ij}(\theta) = -E_\theta \left[\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f \right] \qquad i,j=1,\ldots,k$$

Citamos dos condiciones de regularidad adicionales que nos serán útiles a lo largo de este capítulo:

E.- $\lim\limits_{x\to-\infty} F(x,\theta) \left[\frac{\partial}{\partial\theta_r} \log F(x,\theta) \quad \frac{\partial}{\partial\theta_s} \log F(x,\theta) \right] = 0$

F.- $\lim\limits_{x\to+\infty} \left[1-F(x,\theta) \right] \left[\frac{\partial}{\partial\theta_r} \log(1-F(x,\theta)) \quad \frac{\partial}{\partial\theta_s} \log(1-F(x,\theta)) \right] =$

$$r,s=1,\ldots,k$$

siendo $F(x,\theta) = P_\theta(X \leqslant x)$

Resumimos, a continuación, las propiedades que como medida de información posee la matriz de Fisher.

Sea $\varepsilon_Y = \{(Y,\Omega_Y,\mathcal{B});Q_\theta:\theta \in \Theta\}$ otro experimento con el mismo espacio paramétrico Θ. Usaremos $\underset{\sim}{O}$ para denotar la matriz nula.

Propiedades

(1) $I_X(\theta) \geq 0 \quad \forall\theta \in \Theta$

$I_X(\theta) = \underset{\sim}{0} \iff f(x,\theta)$ no depende de θ

(2) $I_{(X,Y)}(\theta) = I_X(\theta)+I_{Y|X}(\theta) \quad \forall\theta \in \Theta$

$I_{(X,Y)}(\theta) = I_X(\theta)+I_Y(\theta) \quad \forall\theta \in \Theta$, si X e Y son independientes.

(3) $I_X(\theta) \geq I_{T(X)}(\theta) \quad \forall\theta \in \Theta$ y todo estadístico $T(X)$

$I_X(\theta) = I_{T(X)}(\theta) \iff T(X)$ es suficiente para θ

(4) Convexidad

$\alpha \geq 0 \quad \alpha I_X(f_\theta)+(1-\alpha)I_X(g_\theta) \geq I_X(\alpha f_\theta+(1-\alpha)g_\theta) \quad \forall\theta \in \Theta$

para toda $\{f_\theta\}$,$\{g_\theta\} \in \mathcal{F}$, familia de funciones de densidad con el mismo espacio paramétrico Θ.

En esta propiedad ha sido preciso cambiar la notación $I_X(\theta)$ por $I_X(h_\theta)$ para poder expresar dicha propiedad con claridad. $I_X(h_\theta)$ quiere indicar la matriz de Fisher, a partir del experimento ε_X, cuando la densidad de X es h_θ.

(5) Si $\varepsilon_X \gtrsim \varepsilon_Y \Rightarrow I_X(\theta) \geq I_Y(\theta) \quad \forall\theta \in \Theta$

(6) Pérdida de Información debida al Agrupamiento

La secuencia de matrices $\{I_{g_n}(\theta)\}$ converge a la matriz $I_X(\theta)$ elemento a elemento, siendo $\{g_n\}$ una sucesión de particiones tales que g_{n+1} es más fina que g_n para todo n (Apostol; 1986, p. 170). Esto implica que

$\|I_{g_n}-I_X\| \xrightarrow[n\to\infty]{} 0 \quad \forall\theta \in \Theta$, donde $\|\ \|$ denota cualquier norma matricial.

Demostraciones

Las tres primeras propiedades son inmediatas (la primera por la definición de $I_X(\theta)$ y las dos siguientes de forma similar al caso univariante). Para su demostración, ver (Fourgeaud y Fuchs, 1972, pp. 216-220).

La propiedad cuarta fue probada por Stam (1959). La quinta por Goel y DeGroot (1979), bajo condiciones de regularidad estandard como las que aparecen en (Kullback; 1968, pp. 26-27). La demostración de la propiedad sexta aparece en (Ferentinos y Papaioannou, 1979) añadiendo las condiciones E y F a las de regularidad A, B y C de la matriz de información de Fisher. Igualmente prueban (Ferentinos y Papaioannou, 1983) que la relación $\sup_{g \in G} I_g(\theta) = I_X(\theta)$ no es cierta en general para G, conjunto de todas las particiones de Ω_X.

1.3. MEDIDAS DE INFORMACION REALES BASADAS EN LA MATRIZ DE FISHER

1.3.1. INTRODUCCION

A la vista de estas propiedades podemos concluir que la matriz de Fisher posee las principales propiedades de una medida de información y en ello radica su importancia.

Sin embargo, pensemos en el problema de evaluar la posible pérdida de información que, acerca del parámetro, se produce cuando aparece una variable de censura, o aquel otro en que, en tre varios experimentos con el mismo espacio paramétrico deseamos elegir el que proporciona mayor información acerca de dicho parámetro. En éstos y otros muchos problemas nuestro interés será cuantificar la información, y en este sentido, la matriz de Fisher es inoperante y por consiguiente insuficiente para nuestros objetivos.

A esta insuficiencia práctica podemos añadir otra intuitiva de la matriz de Fisher como medida de información. Así, mientras los elementos de la diagonal principal de dicha matriz, $I_X^{ii}(\theta)$ son las medidas de información de Fisher (unidimensionales) acerca de cada parámetro θ_i, dados los valores del resto de parámetros perturbadores θ_j, $j\neq i$; no existe, sin embargo, una interpretación teórica análoga en términos de información para los elementos que están fuera de la diagonal principal. Parece pues, difícil medir la información acerca de un parámetro k-variante por medio de una matriz $k\times k$, algunos de cuyos elementos no tienen connotación de información.

Una forma de solventar estas faltas sería definir una medida real asociada a la matriz de Fisher que tuviera una interpretación intuitiva de medida de información y como tal,buenas propiedades.El siguiente apartado persigue este objetivo.

1.3.2. JUSTIFICACION INTUITIVA

En el caso univariante, $\Theta\subset\mathbb{R}$, la medida de información de Fisher se define, bajo condiciones de regularidad, como:

$$V_\theta\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right] \quad \text{con} \quad E_\theta\left[\frac{\partial}{\partial\theta}\log f(X,\theta)\right] = 0$$

En el caso k-variante, $\Theta \subset \mathbb{R}^k$, su homóloga, la matriz de información de Fisher $I_X(\theta)$ se define, bajo condiciones de regularidad, como la matriz de varianzas y covarianzas de las variables $U_i = \frac{\partial}{\partial \theta_i} \log f(X,\theta)$ $i=1,\dots,k$, con $E_\theta[U_i] = 0$ $i=1,\dots,k$, siendo por tanto una medida de la dispersión multivariante de las variables U_i, $i=1,\dots,k$. En este punto, la información de Fisher es pues sinónimo de dispersión.

Vamos a intentar, sin embargo, reducir a un número la medida de la dispersión multivariante.

Pensemos, para ello, en la transformación de componentes principales de la variable

$$U = \begin{pmatrix} U_1 \\ \vdots \\ U_k \end{pmatrix}$$

es decir: $U \longrightarrow Z = \Gamma'(U - E_\theta(U)) = \Gamma' U$

donde Γ es una matriz ortogonal y Γ' su traspuesta, verificando: $\Gamma' I_X \Gamma = \Lambda$ y Λ es la matriz diagonal con elementos λ_i, autovalores de $I_X(\theta)$ tales que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$

Las variables $Z_1, Z_2, \dots, Z_k$ (componentes de Z) llamadas componentes principales de U, verifican:

$$\left.\begin{array}{l} E(Z_i) = 0 \\ V(Z_i) = \lambda_i \end{array}\right\} i=1,\dots,k$$

$$Cov(Z_i, Z_j) = 0 \quad i \neq j$$

$$V(Z_1) \geq V(Z_2) \geq \dots \geq V(Z_k) \geq 0$$

$$\sum_{i=1}^{k} V(Z_i) = \sum_{i=1}^{k} V(U_i)$$

De esta forma construimos un nuevo sistema de variables que se reparten la varianza total de un modo más heterogéneo, las primeras con mucha varianza y las últimas con poca, con el objetivo de poder resumir la mayor parte de la variabilidad de las U_i, usando únicamente las componentes principales Z_i con - varianzas más altas, reduciendo por consiguiente la dimensión. Puesto que las componentes principales son incorreladas con varianzas $\lambda_1,\ldots,\lambda_k$, parece natural definir la dispersión "global" de las U_i mediante alguna función monótona creciente y simétrica de $\lambda_1,\lambda_2,\ldots,\lambda_k$ (Mardia, Kent y Bibby, 1979).

Este razonamiento nos proporciona una motivación intuitiva para proponer como medidas de información, cuando θ es k-variante, las siguientes funciones reales:

a) $S_X(\theta) = f_1(\lambda_1,\ldots,\lambda_k) = \alpha \sum_{i=1}^{k} \lambda_i$ $\quad\alpha$ constante, $\alpha>0$

b) $D_X(\theta) = f_2(\lambda_1,\ldots,\lambda_k) = \left[\prod_{i=1}^{k} \lambda_i\right]^{\alpha}$ $\quad\alpha$ constante, $\alpha>0$

c) $M_X(\theta) = f_3(\lambda_1,\ldots,\lambda_k) = \left[\sum_{i=1}^{k} \lambda_i^2\right]^{1/2}$

La definición de las dos primeras medidas supone que α es una constante positiva prefijada de antemano. Destacamos por su interpretación los valores $\alpha=1$ y $\alpha=\frac{1}{k}$.

Para $\alpha=1$, $S_X(\theta)$ y $D_X(\theta)$ son la traza y el determinante de la matriz $I_X(\theta)$ respectivamente.

Para $\alpha=\frac{1}{k}$, $S_X(\theta)$ y $D_X(\theta)$ son la media aritmética y geométrica respectivamente, de los autovalores de la matriz $I_X(\theta)$.

Por último, $M_X(\theta)$ es la norma euclídea de la matriz $I_X(\theta)$.

Ferentinos y Papaioannou (1981), basándose en las propiedades que como medidas de información poseen, proponen las siguientes medidas:

a') Traza $[I_X(\theta)]$

b') Determinante $[I_X(\theta)]$

c') λ_i, i-ésimo autovalor ordenado de $I_X(\theta)$, $i=1,\dots,k$

d') $\Omega = \sum_{i=1}^{k} \omega_i\lambda_i \quad \omega_i > 0 \quad i=1,\dots,k$

Lógicamente, estamos de acuerdo con las dos primeras, pero no tanto con las dos últimas. El autovalor i-ésimo no parece - una buena medida de información pues sólo reflejará una parte de la dispersión de $I_X(\theta)$. De elegir algún autovalor, elegiríamos el mayor λ_1.

Respecto a la medida Ω, no es simétrica respecto a los autovalores, salvo cuando $\omega_i=\omega$, $i=1,\dots,k$, que coincide con $S_X(\theta)$ para $\alpha=\omega$. No hay razón para ponderar más unos autovalores que otros.

Para contrastar la validez de las medidas $S_X(\theta)$, $D_X(\theta)$ y $M_X(\theta)$ (propuestas anteriormente) como medidas de información, resumimos sus principales propiedades en el siguiente apartado.

1.3.3. PROPIEDADES DE $S_X(\theta)$, $D_X(\theta)$ y $M_X(\theta)$

Comenzaremos exponiendo las propiedades de los autovalores de la matriz $I_X(\theta)$ pues nos serán de utilidad en algunas demostraciones.

Sean $\lambda_X^1(\theta) \geq \lambda_X^2(\theta) \geq \dots \geq \lambda_X^k \geq 0$ los autovalores de la matriz $I_X(\theta)$.

Propiedades de $\lambda_X^i(\theta)$

(1) $\lambda_X^i(\theta) \geq 0 \quad \forall\theta \in \Theta$

(1') $\lambda_X^i(\theta) = 0 \quad \forall i=1,\ldots,k \iff f(x,\theta)$ no depende de θ

(2) $\lambda_X^i(\theta) \geq \lambda_T^i(\theta) \quad \forall\theta \in \Theta$

(2') $\lambda_X^i(\theta) = \lambda_T^i(\theta) \quad \forall i=1,\ldots,k \iff$ T es suficiente para θ

(3) Convexidad (sólo para el mayor de los autovalores)

$\alpha\geq 0 \qquad \alpha\lambda_X^1(f_\theta)+(1-\alpha)\lambda_X^1(g_\theta) \geq \lambda_X^1(\alpha f_\theta+(1-\alpha)g_\theta) \quad \forall\theta \in \Theta$

y $\{f_\theta\}$, $\{g_\theta\} \in F$

$\lambda_X^1(h_\theta)$ indica el mayor autovalor de la matriz $I_X(h_\theta)$,

$\{h_\theta\} \in F$

(4) Si $\varepsilon_X \gtrsim \varepsilon_Y \Rightarrow \lambda_X^i(\theta) \geq \lambda_Y^i(\theta) \quad \forall\theta \in \Theta$

(5) Pérdida de Información debida al Agrupamiento

$$\sup_{g \in G} \lambda_g^i(\theta) = \lambda_X^i(\theta) \quad \forall\theta \in \Theta$$

$\lambda_g^i(\theta)$ indica el autovalor i-ésimo de la matriz $I_g(\theta)$

Demostraciones

(1) Por ser $I_X(\theta) \geq 0$

(1') Teniendo en cuenta la descomposición espectral de la matriz $I_X(\theta)$, sabemos que $\lambda_X^i(\theta) = 0 \quad \forall i \iff I_X(\theta) = \underset{\sim}{0}$ (matriz nula).

La propiedad (1) de $I_X(\theta)$ completa la demostración.

(2) $I_{T(X)}(\theta)$ e $I_X(\theta)-I_{T(X)}(\theta)$ son simétricas e $I_X(\theta)-I_{T(X)}(\theta) \geq 0$

La propiedad es consecuencia del teorema 3, p. 117 de Bellman (1970).

(2') $\Rightarrow$) $\lambda^i_X(\theta) = \lambda^i_T(\theta) \quad \forall i=1,\dots,k \Rightarrow$ Traza $[I_X(\theta)]$ = = Traza $[I_T(\theta)] \Longleftrightarrow$ Traza $[I_X(\theta)- I_T(\theta)] = 0$.

Ahora bien, la traza de una matriz semidefinida positiva es cero si y sólo si todos sus autovalores son cero y hemos visto, a su vez, que ésto es equivalente a

$I_X(\theta)-I_T(\theta) = \underset{\sim}{0} \Longleftrightarrow I_X(\theta) = I_T(\theta) \Longleftrightarrow$ T suficiente para θ

dándose la última equivalencia por la propiedad (3) de $I_X(\theta)$

$\Leftarrow$) Es inmediata de la propiedad (3) de la matriz $I_X(\theta)$

(3) y (5) en (Ferentinos y Papaioannou; 1983)

(4) en (Ferentinos y Papaioannou; 1982)

□

La propiedad (5) requiere las condiciones de regularidad adicionales E y F, al igual que la propiedad correspondiente de la matriz de Fisher.

1.3.3.1. MEDIDA PARAMETRICA DE INFORMACION $S_X(\theta)$

En el apartado 1.3.2 definíamos la medida paramétrica de información $S_X(\theta)$ mediante la siguiente función real:

$$S_X(\theta) = \alpha \sum_{i=1}^{k} \lambda_X^i(\theta) \quad \text{para un valor de } \alpha > 0 \text{ prefijado}$$

Otras expresiones equivalentes de esta medida serán:

$$S_X(\theta) = \alpha \text{ Traza } [I_X(\theta)] = \alpha \sum_{i=1}^{k} I_X(\theta_i)$$

donde $I_X(\theta_i)$ denota la medida de información de Fisher (unidimensional) acerca del parámetro θ_i, dados los valores del resto de parámetros perturbadores θ_j, $j \neq i$.

Propiedades

(1) $S_X(\theta) \geq 0 \quad \forall \theta \in \Theta$

con la igualdad si y sólo si $f(x,\theta)$ no depende de θ

(2) $S_{X,Y}(\theta) = S_X(\theta) + S_{Y|X}(\theta) \quad \forall \theta \in \Theta$

$S_{X,Y}(\theta) = S_X(\theta) + S_Y(\theta)$ si X e Y son independientes y $\theta \in \Theta$

(3) $S_X(\theta) \geq S_{T(X)}(\theta) \quad \forall \theta \in \Theta$ con la igualdad si y sólo si $T(X)$ es suficiente para θ

(4) Convexidad

$\beta \geq 0 \quad \beta S_X(f_\theta) + (1-\beta) S_X(g_\theta) \geq S_X(\beta f_\theta + (1-\beta) g_\theta) \quad \forall \theta \in \Theta$

y $\{f_\theta\}, \{g_\theta\} \in \mathcal{F}$

siendo $S_X(h_\theta) = \alpha$ Traza $[I_X(h_\theta)] \quad \{h_\theta\} \in \mathcal{F}$

(5) $\varepsilon_X \gtrsim \varepsilon_Y \Rightarrow S_X(\theta) \geq S_Y(\theta) \quad \forall \theta \in \Theta$

(6) Pérdida de Información debida al Agrupamiento

$$\sup_{g \in G} S_g(\theta) = S_X(\theta) \quad \forall \theta \in \Theta \quad \text{con } S_g(\theta) = \alpha \text{ Traza } [I_g(\theta)]$$

Demostraciones

La propiedad (2) es consecuencia inmediata de la propiedad (2) de la matriz de Fisher. El resto de las propiedades puede obtenerse inmediatamente de las propiedades correspondientes de la traza de la matriz de Fisher, basta multiplicar por $\alpha>0$ en los dos miembros de las desigualdades respectivas y tener en cuenta que:

$\sup_t (\alpha f(t)) = \alpha \sup_t f(t)$ si $\alpha \geq 0$ y $f(t)$ es una función real acotada.

Las demostraciones correspondientes a la traza pueden encontrarse en (Ferentinos y Papaioannou, 1979; 1981; 1982; 1983).

De nuevo, la propiedad (6) de pérdida de información debida al agrupamiento requiere que la matriz $I_X(\theta)$ verifique las condiciones de regularidad adicionales E y F.

1.3.3.2. MEDIDA PARAMETRICA DE INFORMACION $D_X(\theta)$

Denotábamos por $D_X(\theta)$ a la medida paramétrica de información definida mediante la siguiente función real:

$D_X(\theta) = \left[\prod_{i=1}^{k} \lambda_X^i(\theta) \right]^\alpha$ que podemos expresar alternativamente como:

$D_X(\theta) = \left[|I_X(\theta)| \right]^\alpha$ para un valor de $\alpha>0$ prefijado, donde $|\ |$ denota el determinante de una matriz.

Propiedades

(1) $D_X(\theta) \geqslant 0 \quad \forall\theta \in \Theta$

Si $f(x,\theta)$ no depende de $\theta \Rightarrow D_X(\theta) = 0$

El recíproco no se da, ya que una matriz semidefinida positiva puede ser singular.

(2) Para $\alpha = \frac{1}{k}$, $k > 1$; si $I_X(\theta)$, $I_{Y|X}(\theta)$ e $I_Y(\theta)$ son definidas positivas

$$D_{X,Y}(\theta) \geqslant D_X(\theta)+D_{Y|X}(\theta) \quad \forall\theta \in \Theta$$

$D_{X,Y}(\theta) \geqslant D_X(\theta)+D_Y(\theta) \quad \forall\theta \in \Theta$ si X e Y son independientes.

(3) Multiplicidad

Sean $X_1, X_2, \ldots, X_n$ n observaciones independientes de la variable X

$$D_{X_1,X_2,\ldots,X_n}(\theta) = n^{k\alpha} D_X(\theta) \quad \forall\theta \in \Theta$$

Para $\alpha = \frac{1}{k}$ $\quad D_{X_1,X_2,\ldots,X_n}(\theta) = nD_X(\theta) \quad \forall\theta \in \Theta$ (Aditividad)

(4) $D_X(\theta) \geqslant D_T(\theta) \quad \forall\theta \in \Theta$

Si T es suficiente para $\theta \Rightarrow D_X(\theta) = D_T(\theta)$

(5) $\varepsilon_X \gtrsim \varepsilon_Y \Rightarrow D_X(\theta) \geqslant D_Y(\theta) \quad \forall\theta \in \Theta$

(6) Pérdida de información debida al agrupamiento

$$\sup_{g \in G} D_g(\theta) = D_X(\theta) \quad \forall\theta \in \Theta \qquad \text{con } D_g(\theta) = \left[|I_g(\theta)|\right]^\alpha$$

Demostraciones

La propiedad (2) es consecuencia del resultado: $|A+B|^{1/k} \geq |A|^{1/k}+|B|^{1/k}$ para $A>0$, $B>0$, matrices reales de orden k (Rao; 1973, p. 70).

Las restantes propiedades pueden obtenerse fácilmente de las correspondientes al determinante de la matriz de Fisher, teniendo en cuenta que $\alpha>0$ y que $\sup(f^{\alpha}) = (\sup f)^{\alpha}$ si $\alpha\geq 0$ y f es una función real acotada, $f\geq 0$.

Las demostraciones de las propiedades análogas para el determinante pueden encontrarse: (1) y (4) en (Ferentinos y Papaioannou; 1981); (5) en (Goel y DeGroot; 1979); (6) en (Ferentinos y Papaioannou; 1983). La propiedad (3) se obtiene de forma inmediata a partir de la propiedad (2) de la matriz de Fisher.

Otra vez, la propiedad (6) requiere que la matriz $I_X(\theta)$ verifique las condiciones de regularidad adicionales E y F.

1.3.3.3. MEDIDA PARAMETRICA DE INFORMACION $M_X(\theta)$

Antes de entrar en la definición y propiedades de la medida $M_X(\theta)$, daremos un conjunto de resultados acerca de la norma euclídea de una matriz que nos servirá de apoyo en el estudio de las propiedades de $M_X(\theta)$.

Definición

La norma euclídea de una matriz cuadrada $A=(a_{ij})$ $i,j=1,\dots,k$ es el número no negativo $\|A\|$ obtenido de la expresión

$$\|A\| = \left[\sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}^2\right]^{1/2} \qquad \text{(Rao; 1973, p. 63)}$$

Claramente:

$||A|| = (\text{Traza } A'A)^{1/2} = (\text{Traza } AA')^{1/2}$. Si además A es simétrica y $A = \Gamma\Lambda\Gamma'$ es su descomposición espectral con Λ matriz diagonal de los autovalores λ_i de A, entonces:

$$A'A = \Gamma\Lambda\Gamma'\Gamma\Lambda\Gamma' = \Gamma\Lambda^2\Gamma' \text{ de donde:}$$

$$||A|| = \left[\sum_{i=1}^{k} \lambda_i^2\right]^{1/2}$$

Como norma matricial verifica las condiciones

1.- $||A|| > 0$ si $A \neq \underset{\sim}{O}$ y $||\underset{\sim}{O}|| = 0$ (1.3.1)

2.- $||cA|| = |c|\ ||A||$ $c \in \mathbb{R}$ (1.3.2)

3.- $||A+B|| \leq ||A|| + ||B||$ (1.3.3)

4.- $||AB|| \leq ||A||\ ||B||$

Además se dan las condiciones de convergencia siguientes:

$$||A^{(k)}-A|| \xrightarrow[k\to\infty]{} 0 \iff A^{(k)} \xrightarrow[k\to\infty]{} A \quad \text{(elemento a elemento)}$$

$$\text{y} \quad A^{(k)} \xrightarrow[k\to\infty]{} A \implies ||A^{(k)}|| \longrightarrow ||A|| \qquad (1.3.4)$$

Todas estas condiciones pueden encontrarse en (Faddeev-Faddeeva, 1963).

Estamos ya en condiciones de analizar la medida $M_X(\theta)$ introducida en el apartado 1.3.2. Denotábamos por $M_X(\theta)$ a la medida paramétrica de información definida mediante la siguiente función real:

$M_X(\theta) = \left[\sum_{i=1}^{k} (\lambda_X^i(\theta))^2\right]^{1/2}$. Una expresión alternativa será

$M_X(\theta) = ||I_X(\theta)||$

Propiedades

(1) $M_X(\theta) \geq 0 \quad \forall\theta \in \Theta$

con la igualdad si y sólo si $f(x,\theta)$ no depende de θ

(2) $M_{X,Y}(\theta) \leq M_X(\theta)+M_{Y|X}(\theta) \quad \forall\theta \in \Theta$

$M_{X,Y}(\theta) \leq M_X(\theta)+M_Y(\theta)$ si X e Y son independientes y $\theta \in \Theta$

(3) Sean $X_1,X_2,\dots,X_n$ n observaciones independientes de la variable X

$M_{X_1,X_2,\dots X_n}(\theta) = nM_X(\theta) \quad \forall\theta \in \Theta$

(4) $M_X(\theta) \geq M_{T(X)}(\theta) \quad \forall\theta \in \Theta$ con la igualdad si y sólo si $T(X)$ es suficiente para θ

(5) Convexidad

$\alpha\geq 0 \quad \alpha M_X(f_\theta)+(1-\alpha)M_X(g_\theta) \geq M_X(\alpha f_\theta+(1-\alpha)g_\theta) \quad \forall\theta\in \Theta$

y $\{f_\theta\}$, $\{g_\theta\} \in \mathcal{F}$

(6) $\varepsilon_X \gtrsim \varepsilon_Y \Rightarrow M_X(\theta) \geq M_Y(\theta) \quad \forall\theta \in \Theta$

(7) Pérdida de Información debida al Agrupamiento

$\sup_{g \in G} M_g(\theta) = M_X(\theta) \quad \forall\theta\in \Theta$ con $M_g(\theta) = ||I_g(\theta)||$

Demostraciones

Usaremos la notación $\lambda_i(A)$ para indicar el autovalor i-ésimo de la matriz A.

(1) Es consecuencia de (1.3.1) y de la propiedad (1) de la matriz $I_X(\theta)$. También puede obtenerse de las propiedades (1) y (1') de los autovalores $\lambda_X^i(\theta)$.

(2) Consecuencia de (1.3.3) y de la propiedad (2) de la matriz $I_X(\theta)$

(3) Teniendo en cuenta que

$$I_{X_1,X_2,\ldots,X_n}(\theta) = nI_X(\theta) \quad \text{(propiedad (2) de aditividad de la matriz } I_X(\theta))$$

entonces

$$\lambda_i(I_{X_1,X_2,\ldots,X_n}(\theta)) = n\lambda_i(I_X(\theta)) \quad i=1,\ldots,n$$

de esta forma $\|I_{X_1,\ldots,X_n}(\theta)\| = n\|I_X(\theta)\|$

(4) Consecuencia de las propiedades (2) y (2') de los autovalores $\lambda_X^i(\theta)$

(5) Por la propiedad (4) de convexidad de $I_X(\theta)$

$\alpha\geq 0 \quad I_X(\alpha f_\theta+(1-\alpha)g_\theta) \leq \alpha I_X(f_\theta)+(1-\alpha)I_X(g_\theta)$

Estamos en las condiciones del teorema 3, p. 117 de Bellman (1970) y por tanto:

$$\lambda_i(I_X(\alpha f_\theta+(1-\alpha)g_\theta)) \leq \lambda_i(\alpha I_X(f_\theta)+(1-\alpha)I_X(g_\theta)) \quad i=1,\ldots,k$$

Por ser los autovalores $\lambda_i(I_X(\alpha f_\theta+(1-\alpha)g_\theta)) \geq 0 \quad i=1,\ldots,k$ entonces:

$$\sum_{i=1}^{k} \lambda_i^2(I_X(\alpha f_\theta+(1-\alpha)g_\theta)) \leq \sum_{i=1}^{k} \lambda_i^2(\alpha I_X(f_\theta)+(1-\alpha)I_X(g_\theta))$$

es decir:

$$\| I_X(\alpha f_\theta+(1-\alpha)g_\theta)\| \leq \|\alpha I_X(f_\theta)+(1-\alpha)I_X(g_\theta)\| \qquad (1.3.5)$$

aplicando a la norma del segundo miembro las condiciones (1.3.3) y (1.3.2) respectivamente, obtenemos el resultado deseado.

(6) Es inmediata de la propiedad (4) de los autovalores $\lambda_X^i(\theta)$

$\square$

Una vía alternativa de demostración de las propiedades (4) y (6) y de la condición (1.3.5) es utilizar un lema de Okamoto y Kanazawa (1968) que exponemos en el apéndice A.1 al final de este trabajo.

(7) Suponemos, como ya es habitual en esta propiedad, que la matriz $I_X(\theta)$ verifica las condiciones de regularidad adicionales E y F.
Antes de demostrar la propiedad enunciamos un teorema previo.

<u>Teorema</u>

En las condiciones de regularidad A,B,C de la matriz de Fisher, junto a las adicionales E y F; si $\{g_n\}$ es una sucesión de particiones, tales que g_{n+1} es más fina que g_n para todo n, entonces

$$\lim_{n\to\infty} \| I_{g_n}(\theta)\| = \| I_X(\theta)\| \quad \forall\theta \in \Theta$$

<u>Demostración</u>

Ferentinos y Papainoannou (1979) muestran en las condiciones del teorema que:

$$I_{g_n}(\theta) \xrightarrow[n\to\infty]{} I_X(\theta) \quad \text{(elemento a elemento)}$$

La demostración se concluye aplicando la condición (1.3.4)

$\square$

Estamos ya en condiciones de demostrar la propiedad (7)

Sea $\{g_n\}$ una sucesión de particiones como la definida en el teorema anterior.

Por la propiedad (4) de información máxima sabemos que:

$$\| I_{g_n}(\theta) \| \leqslant \| I_X(\theta) \| \quad \forall n \ \forall \theta \in \Theta$$

Por el teorema anterior

$$\lim_{n \to \infty} \| I_{g_n}(\theta) \| = \| I_X(\theta) \| \quad \forall \theta \in \Theta$$

Entonces $$\sup_n \| I_{g_n}(\theta) \| = \| I_X(\theta) \| \quad \forall \theta \in \Theta$$

Dado que $g_n \in G \quad \forall n$, tenemos

$$\sup_{g \in G} \| I_g(\theta) \| \geqslant \sup_n \| I_{g_n}(\theta) \| = \| I_X(\theta) \| \qquad \forall \theta \in \Theta \qquad (1.3.6)$$

De nuevo, la propiedad (4) implica que

$$\| I_g(\theta) \| \leqslant \| I_X(\theta) \| \quad \text{para toda } g \in G \ \text{ y } \ \theta \in \Theta$$

Por consiguiente

$$\sup_{g \in G} \| I_g(\theta) \| \leqslant \| I_X(\theta) \| \qquad \forall \theta \in \Theta \qquad (1.3.7)$$

Combinando las relaciones (1.3.6) y (1.3.7) se obtiene el resultado.

□

Como resumen del análisis de las propiedades de las medidas de información $S_X(\theta)$, $D_X(\theta)$ y $M_X(\theta)$ podemos concluir que tanto $S_X(\theta)$ como $M_X(\theta)$ poseen las principales propiedades de una medida de información, y por ésto serán preferibles a la medida $D_X(\theta)$. Esta última no verifica, más que parcialmente, las propiedades de no negatividad y de invariancia para transformaciones suficientes; asímismo no cumple la propiedad de convexidad. Si bien esta última propiedad no descartaría a $D_X(\theta)$ como medida de información, si lo harían las dos anteriores, que constituyen, junto a la de información máxima, un mínimo de propiedades a exigir a toda medida de información.

Sin embargo, si la matriz de Fisher $I_X(\theta)$ es no singular $\forall\ \theta \in \Theta$ se verifican claramente ambas propiedades.

En caso contrario, basada en $D_X(\theta)$ podemos definir una medida de información alternativa, que denotamos por $D^*_X(\theta)$ y cuya expresión es:

$$D^*_X(\theta) = \left[|I+I_X(\theta)|^{\alpha} \right] -1 = \left[\prod_{i=1}^{k} (1+\lambda^i_X(\theta)) \right]^{\alpha} -1$$

para un valor de $\alpha>0$ prefijado, siendo I la matriz identidad.

A partir de las propiedades de $\lambda^i_X(\theta)$ es sencillo comprobar que esta medida verifica las propiedades de no negatividad, información máxima e invariancia para transformaciones suficientes. Igualmente verifica las propiedades de suficiencia de experimentos y de pérdida de información debida al agrupamiento.

A continuación, nos centraremos en el estudio del comportamiento de las medidas paramétricas de información $S_X(\theta)$, $D_X(\theta)$ y $M_X(\theta)$ respecto de transformaciones biyectivas del espacio paramétrico.

1.3.4. TRANSFORMACIONES BIYECTIVAS DEL ESPACIO PARAMETRICO

Sea $\phi=f(\theta)$ una transformación biyectiva de θ, de forma que $\theta=g(\phi)$ es la transformación inversa de f. Sea $\Psi=f(\Theta)$ el espacio paramétrico transformado.

Llamamos $A=(a_{ij})$ a la matriz cuadrada, cuyo elemento (i,j) es $a_{ij} = \frac{\partial \theta_i}{\partial \phi_j}$, i,j=1,...,k. Siguiendo la notación empleada,

$I_X(\phi)$ representará la matriz de información de Fisher, acerca de ϕ, según la variable X. Análogamente denotaremos $S_X(\phi)$, $D_X(\phi)$ y $M_X(\phi)$ a las medidas correspondientes acerca de ϕ.

Enunciamos primeramente un resultado conocido para la matriz de Fisher (Fisher; 1956, p. 155) y a continuación un lema relativo a los autovalores, ambos de gran utilidad en adelante.

Proposición

$$I_X(\phi) = A'I_X(\theta)A \qquad (1.3.8)$$

Este resultado proporciona la expresión que relaciona las matrices de Fisher, según X, acerca de los parámetros θ y ϕ.

Análogamente a (1.3.8) podríamos escribir:

$$I_X(\theta) = B'I_X(\phi)B \qquad (1.3.9)$$

siendo $B=(b_{ij})$ la matriz cuyo elemento (i,j) es $b_{ij}=\frac{\partial\phi_i}{\partial\theta_j}$ $i,j=1,\dots,k$ de forma que si A es no singular $B=A^{-1}$.

Con objeto de clarificar la notación, en lo sucesivo para cualquier matriz $H\geqslant 0$ escribiremos $\lambda_i(H)$ para denotar el autovalor i-ésimo de H, es decir $\lambda_1(H) \geqslant \lambda_2(H) \geqslant \dots \geqslant \lambda_k(H) \geqslant 0$.

Lema

Si $\phi=f(\theta)$ es una transformación lineal ortogonal de θ, entonces

$$\lambda_i(I_X(\phi)) = \lambda_i(I_X(\theta)) \quad i=1,\dots,k \qquad (1.3.10)$$

Demostración

Escribimos en notación matricial $\phi' = (\phi_1,\dots,\phi_k)$ y $\theta' = (\theta_1,\dots,\theta_k)$.

Por ser ϕ una transformación lineal ortogonal de θ, podemos escribir

$\phi=H\theta$ y H es una matriz ortogonal de orden k

$(H'H=HH'=I)$

Es sencillo probar que $H'=A$. Por consiguiente

$I_X(\phi) = HI_X(\theta)H' = HI_X(\theta)H^{-1}$ por lo que las matrices $I_X(\phi)$ e $I_X(\theta)$ tienen el mismo polinomio característico (Rao; 1973, p. 76 prob. 25) y por lo tanto los mismos autovalores.

□

1.3.4.1. MEDIDA DE INFORMACION S_X

Daremos una serie de relaciones entre las medidas de información $S_X(\theta)$ y $S_X(\phi)$

Propiedad 1

Si la matriz A es no singular, entonces

$$S_X(\phi) \leq \lambda_1(AA')S_X(\theta)$$

$$S_X(\phi) \geq \lambda_k(AA')S_X(\theta)$$

siendo $\lambda_1>0$ y $\lambda_k>0$.

Demostración

$I_X(\theta)\geq 0$ y $AA'>0$ por ser A no singular (Faddeev-Faddeeva; 1963, p. 24 teorema 1.3)

Entonces, aplicando el corolario 2.2.1 de (Anderson y Das Gupta; 1963), resulta

$$\lambda_k(AA')\lambda_i(I_X(\theta)) \leq \lambda_i(I_X(\theta)AA') \leq \lambda_1(AA')\lambda_i(I_X(\theta))$$
$$i=1,\ldots,k$$

Sumando en i

$$\lambda_k(AA') \text{ Traza } (I_X(\theta)) \leq \text{Traza } (I_X(\theta)AA') \leq$$
$$\leq \lambda_1(AA') \text{ Traza } (I_X(\theta)) \qquad (1.3.11)$$

Por otra parte (1.3.8) implica que

$$\text{Traza } (I_X(\phi)) = \text{Traza } (A'I_X(\theta)A) = \text{Traza } (I_X(\theta)AA') \qquad (1.3.12)$$

con la última igualdad por el hecho de que Traza (GH)=Traza (HG)

Teniendo en cuenta que $S_X(\tau) = \alpha$ Traza $(I_X(\tau))$ las relaciones (1.3.11) y (1.3.12) nos conducen al resultado.

Por último, por ser AA'>0, todos sus autovalores son positivos.

□

Corolario 1

Si A es simétrica y no singular, entonces

$$S_X(\phi) \leq \lambda_1^2(A)S_X(\theta)$$
$$S_X(\phi) \geq \lambda_k^2(A)S_X(\theta)$$

Demostración

Es inmediata de la propiedad 1

□

Propiedad 2

Si $I_X(\theta)$ es definida positiva, entonces

$$S_X(\phi) \leq \alpha\lambda_1(I_X(\theta))\|A\|^2$$
$$S_X(\phi) \geq \alpha\lambda_k(I_X(\theta))\|A\|^2$$

con $\lambda_1 > 0$ y $\lambda_k > 0$.

Demostración

Sabemos que Traza($A' I_X(\theta) A$) = Traza($AA' I_X(\theta)$)

Ahora $I_X(\theta) > 0$ y $AA' \geq 0$ para cualquier matriz A (Mardia, Kent y Bibby;1979, p. 476).

Intercambiando ambas matrices en la demostración de la propiedad 1 y teniendo en cuenta que $\| A \|^2$ = Traza(AA'), llegamos al resultado.

$\square$

Propiedad 3

Si $\phi = f(\theta)$ es una transformación lineal ortogonal de θ entonces

$$S_X(\phi) = S_X(\theta)$$

Demostración

Es consecuencia inmediata de (1.3.10).

$\square$

Podemos concluir pues, que la medida de información S_X es invariante para transformaciones lineales ortogonales del espacio paramétrico.

Observación

Para transformaciones biyectivas, no lineales ortogonales, en general no se da la invariancia de la medida S_X.

1.3.4.2. MEDIDA DE INFORMACION D_X

Ocuparemos este epígrafe en dar una serie de relaciones entre las medidas de información $D_X(\theta)$ y $D_X(\phi)$.

<u>Propiedad 4</u>

$$D_X(\phi) = \left[|A|^2\right]^{\alpha} D_X(\theta)$$

<u>Demostración</u>

Es inmediata de (1.3.8) ya que $|I_X(\phi)| = |A'I_X(\theta)A| = |A'||I_X(\theta)||A|$ por ser todos los factores matrices cuadradas.

Por tanto

$$|I_X(\phi)| = |A|^2|I_X(\theta)|$$

Elevando a α ambos miembros obtenemos el resultado.

□

<u>Corolario 2</u>

Si A es singular para algún $\phi^* \in \Psi$, entonces

$$D_X(\phi^*) = 0$$

<u>Demostración</u>

Es consecuencia inmediata de la propiedad 4.

□

<u>Corolario 3</u>

$$D_X(\phi) = D_X(\theta) \Longleftrightarrow |A| = \pm 1$$

Demostración

Obvia, a partir de la propiedad 4.

□

Corolario 4

Si $\phi=f(\theta)$ es una transformación lineal ortogonal de θ, entonces

$$D_X(\phi) = D_X(\theta)$$

Demostración

Como consecuencia de (1.3.10) $|I_X(\phi)| = |I_X(\theta)|$ que conduce obviamente al resultado.

Otro camino sería comprobar que estamos en las condiciones del corolario 3 lo que resulta igualmente sencillo ya que, por ser ϕ lineal ortogonal $A=H' \Rightarrow |A|=|H|=\pm 1$

□

Es decir, la medida de información D_X es invariante para transformaciones biyectivas del espacio paramétrico tales que $|A|=\pm 1$ y entre éstas para las transformaciones lineales ortogonales.

1.3.4.3. MEDIDA DE INFORMACION M_X

Vamos a dar dos propiedades acerca de la relación entre las medidas de información $M_X(\theta)$ y $M_X(\phi)$.

Propiedad 5

Si A es no singular, entonces

$$M_X(\phi) \leq \lambda_1(AA')M_X(\theta)$$

$$M_X(\phi) \geq \lambda_k(AA')M_X(\theta)$$

con $\lambda_1>0$ y $\lambda_k>0$

Demostración

$I_X(\theta) \geq 0$ y $AA' > 0$ por hipótesis. Entonces

$$\lambda_k(AA')\lambda_i(I_X(\theta)) \leq \lambda_i(I_X(\theta)AA') \leq \lambda_i(I_X(\theta))\lambda_1(AA')$$

$$i=1,\ldots,k$$

(Anderson y Das Gupta; 1963)

Por ser positivos los tres miembros de las desigualdades anteriores, podemos escribir

$$\lambda_k^2(AA')\lambda_i^2(I_X(\theta)) \leq \lambda_i^2(I_X(\theta)AA') \leq \lambda_i^2(I_X(\theta))\lambda_1^2(AA') \quad i=1,\ldots,k \tag{1.3.13}$$

pero $\lambda_i(A'I_X(\theta)A) = \lambda_i(I_X(\theta)AA')$ (Mardia, Kent y Bibby; 1979, p. 468 th. A.6.2)

$i=1,\ldots,k$

Por otra parte

$$||I_X(\phi)||^2 = ||A'I_X(\theta)A||^2 = \sum_{i=1}^{k} \lambda_i^2(A'I_X(\theta)A)$$

Entonces, a partir de (1.3.13) obtenemos

$$\lambda_k^2(AA') \sum_{i=1}^{k} \lambda_i^2(I_X(\theta)) \leq ||I_X(\phi)||^2 \leq \lambda_1^2(AA') \sum_{i=1}^{k} \lambda_i^2(I_X(\theta))$$

El resultado se obtiene, teniendo en cuenta que

$$\sum_{i=1}^{k} \lambda_i^2(I_X(\theta)) = ||I_X(\theta)||^2$$

□

Igual que con la medida S_X, si $I_X>0$ podemos dar otras acotaciones de menor interés para $M_X(\phi)$.

Propiedad 6

Si $\phi=f(\theta)$ es una transformación lineal ortogonal de θ, entonces:

$$M_X(\phi) = M_X(\theta)$$

Demostración

Es inmediata de (1.3.10), o bien como corolario de la propiedad 5, pues en este caso A es una matriz ortogonal, por lo que $AA'=I$ y $\lambda_i(I)=1 \quad i=1,\dots,k$

$\square$

Por tanto, la medida de información M_X es invariante para transformaciones lineales ortogonales del espacio paramétrico. Igual que ocurría con S_X, para transformaciones biyectivas cualesquiera no se da, en general, la propiedad de invariancia de la medida M_X.

Para concluir este capítulo daremos una propiedad común para las tres medidas de información S_X, D_X y M_X, relativa a la comparación de experimentos.

1.3.5. OTROS CRITERIOS PARA LA COMPARACION DE EXPERIMENTOS

Estudiamos en este apartado, el comportamiento de las medidas de información S_X, D_X y M_X respecto a dos criterios de comparación de experimentos más débiles que el de Blackwell.

Sea $\theta \in \Theta \subset \mathbb{R}^k$ un parámetro k-dimensional, ε_X y ε_Y dos experimentos estadísticos con espacio paramétrico común Θ e $I_X(\theta)$, $I_Y(\theta)$ las matrices de información de Fisher para los experimentos ε_X y ε_Y respectivamente, bajo las condiciones de re gularidad estandar. $\varepsilon_X \gtrsim \varepsilon_Y$ indicará que el experimento ε_X es suficiente para el experimento ε_Y, según la definición de Blackwell (1951, 1953).

Definición 1

Decimos que ε_X es <u>suficiente por pares</u> para el experimento ε_Y y denotamos $\varepsilon_X \gtrsim_2 \varepsilon_Y$ si para todo par de valores $\theta^1, \theta^2 \in \Theta$ $\varepsilon_X \gtrsim \varepsilon_Y$, cuando limitamos el espacio paramétrico a contener únicamente los dos valores θ^1 y θ^2.

Claramente $\varepsilon_X \gtrsim \varepsilon_Y \Rightarrow \varepsilon_X \gtrsim_2 \varepsilon_Y$

El recíproco no se da necesariamente, podemos encontrar un contraejemplo en (Bayarri y DeGroot; 1986).

Definición 2

Decimos que ε_X es <u>preferido</u> a ε_Y y denotamos por $\varepsilon_X \gtrsim_F \varepsilon_Y$ siempre que $I_X(\theta)-I_Y(\theta)$ sea semidefinida positiva $\forall \theta \in \Theta$, es decir

$$\varepsilon_X \gtrsim_F \varepsilon_Y \iff I_X(\theta)-I_Y(\theta) \geq 0 \quad \forall \theta \in \Theta$$

Por la propiedad (5) de la matriz de Fisher

$$\varepsilon_X \gtrsim \varepsilon_Y \Rightarrow \varepsilon_X \gtrsim_F \varepsilon_Y$$

Un contraejemplo de que el recíproco no es cierto necesariamente puede encontrarse en (Hansen y Torgersen; 1974).

Además $\varepsilon_X \gtrsim_2 \varepsilon_Y \Rightarrow \varepsilon_X \gtrsim_F \varepsilon_Y$ (Goel y DeGroot; 1979; Bayarri y DeGroot; 1986).

Propiedad 7

Si $\varepsilon_X \gtrsim_F \varepsilon_Y$ se verifican

$$\left.\begin{array}{ll} \text{i)} & S_X(\theta) \geq S_Y(\theta) \\ \text{ii)} & D_X(\theta) \geq D_Y(\theta) \\ \text{iii)} & M_X(\theta) \geq M_Y(\theta) \end{array}\right\} \forall\theta \in \Theta$$

Demostración

Por la definición de $\gtrsim_F$ la matriz $I_X(\theta)-I_Y(\theta) \geq 0 \quad \forall\theta \in \Theta$

Las matrices $I_Y(\theta)$ e $I_X(\theta)-I_Y(\theta)$ son simétricas e $I_X(\theta)-I_Y(\theta) \geq 0 \quad \forall\theta \in \Theta$

Entonces por el teorema 3, p. 117 de (Bellman; 1970) resulta que

$$\lambda_i(I_X(\theta)) \geq \lambda_i(I_Y(\theta)) \quad i=1,\ldots,k \quad \forall\theta \in \Theta$$

Por ser $S_X(\theta)$, $D_X(\theta)$ y $M_X(\theta)$ tres funciones de los autovalores estrictamente crecientes en cada argumento, la propiedad queda demostrada de manera inmediata en sus tres apartados.

$\square$

Esta propiedad puede obtenerse como consecuencia inmediata del lema de Okamoto y Kanazawa (1968) (Apéndice A.1).

Corolario 5

Si $\varepsilon_X \gtrsim_2 \varepsilon_Y$ se verifican

$$\left.\begin{array}{ll} \text{i)} & S_X(\theta) \geq S_Y(\theta) \\ \text{ii)} & D_X(\theta) \geq D_Y(\theta) \\ \text{iii)} & M_X(\theta) \geq M_Y(\theta) \end{array}\right\} \forall\theta \in \Theta$$

Demostración

$\varepsilon_X \gtrsim_2 \varepsilon_Y \implies \varepsilon_X \gtrsim_F \varepsilon_Y$. La propiedad 7 nos conduce al resultado.

$\square$

CAPITULO II

PERDIDA DE INFORMACION EN MODELOS DE SUPERVIVENCIA CENSURADOS

CAPITULO II

PERDIDA DE INFORMACION EN MODELOS DE SUPERVIVENCIA CENSURADOS

2.0. SUMARIO

Como indica el título, este capítulo se dedicará al problema de la pérdida de información que se produce en estudios de supervivencia y fiabilidad cuando las observaciones sobre los tiempos de vida se censuran.

Tras una introducción acerca de los modelos de supervivencia censurados, recogemos algunas recomendaciones en torno a las medidas de información adecuadas para este tipo de modelos (Sección 2.1).

Estamos interesados en evaluar esta pérdida de información originada por la censura, cuando la medida de información adecuada es la matriz de Fisher.

Nuestro punto de partida lo constituirán las medidas reales de dicha pérdida, utilizadas por Brooks (1982) en un contexto bayesiano cuando la medida de información es la medida de Shannon (Sección 2.2).

Sugerimos dos maneras de abordar nuestro objetivo. La primera, desarrollada a lo largo de las secciones 2.3 y 2.4, consiste en definir dos medidas matriciales de la pérdida de información que llamaremos matriz de pérdidas y matriz de eficiencias que justificaremos de manera formal por las propiedades de sus autovalores y de manera intuitiva por el claro significado de dichos autovalores al diagonalizar la matriz de Fisher mediante una transformación biyectiva del parámetro de interés.

Proponemos asímismo tres medidas reales de la pérdida de información, que definimos mediante otras tantas funciones reales de las matrices anteriores y en particular de sus autovalores, heredando de esta forma las buenas propiedades de éstos, destacando la propiedad de invariancia para transformaciones biyectivas del parámetro. Todo este proceso requiere que la matriz de Fisher del modelo no censurado sea no singular.

La segunda forma de abordar nuestro objetivo, menos original que la anterior y que desarrollamos en la sección 2.5, consiste en seleccionar como medidas de información, las medidas reales S_X, D_X y M_X (basadas en la matriz de Fisher) que proponíamos en el capítulo I, y trasladar las definiciones de la pérdida de información, dadas por Brooks, a un contexto no bayesiano.

Al igual que antes, se analizan las propiedades de estas medidas de la pérdida de información, destacando lo siguiente:

Si las medidas de información elegidas son S_X o M_X, las medidas correspondientes de la pérdida de información no son invariantes para transformaciones biyectivas del parámetro, en cambio, la definición de las mismas no requiere la hipótesis restrictiva de no singularidad de la matriz de Fisher correspondiente al modelo no censurado.

En caso de elegir la medida de información D_X, las propias limitaciones de dicha medida exigen la no singularidad de las matrices de Fisher relativas a los modelos no censurado y censurado, a cambio, las medidas de la pérdida de información correspondientes son invariantes para transformaciones biyectivas del parámetro.

Todo este proceso puede sintetizarse en el siguiente esquema:

$$
\begin{tikzcd}
 & \text{Pérdida Matricial de Información} \arrow[dr] & \\
\text{Matriz de Información de Fisher} \arrow[ur] \arrow[dr] & & \text{Medida Real de Pérdida de Información} \\
 & \text{Medida Real de Información} \arrow[ur] &
\end{tikzcd}
$$

2.1. INTRODUCCION

El análisis de supervivencia es un término estadístico que abarca una gran variedad de técnicas estadísticas para el análisis de variables aleatorias positivas.

Generalmente la variable aleatoria es el tiempo hasta el primer fallo (avería) de una componente física (mecánica o eléctrica) o el tiempo hasta la muerte (tiempo de vida) de una unidad biológica (paciente, animal, célula, etc.).

Debido a su gran aplicación en la industria en estudios de fiabilidad, así como en la medicina en numerosos ensayos clínicos, el análisis de supervivencia ha experimentado un fuerte auge en los últimos veinte años, existiendo en la actualidad numerosos textos dedicados exclusivamente a este tema, que incluyen tanto los modelos paramétricos como los no paramétricos. Entre ellos destacamos Kalbfleisch y Prentice (1980), Lee (1980), Miller (1981), Nelson (1982), Cox y Oakes (1984).

Típicamente, el objetivo del análisis de supervivencia es hacer inferencias acerca del parámetro desconocido cuando la distribución de la variable aleatoria pertenece a una familia paramétrica, o bien acerca de algún funcional de la función de distribución, si el modelo para dicha variable es no paramétrico.

Como vehículo para este proceso inferencial, el estadístico dispone de un experimento en el que observa la variable tiempo de vida en una muestra de n unidades.

La primera dificultad que nos vamos a encontrar en la mayor parte de los estudios sobre supervivencia es la imposibili dad de disponer de los valores exactos de la variable para todas las unidades muestrales debido a lo que, de modo genérico, llamamos la censura.

Por ejemplo, en estudios de fiabilidad sobre la duración de una determinada componente, no será útil esperar a la destrucción de la misma. En supervivencia algunos pacientes pueden sobrevivir al final del ensayo clínico, o pasar a otro tratamiento, o morir por otra causa ajena a la enfermedad en estudio produciéndose, en cualquier caso, una observación incompleta para algunas unidades muestrales. Diremos que estas unidades están censuradas.

La formalización de este hecho es la siguiente.

Denotaremos por T a la variable aleatoria tiempo de vida. En presencia de censura, lo que realmente observa el estadístico para cada unidad muestral es la variable aleatoria bidimensional

(X,δ) definida por $X = \min(T,C)$ y $\delta = I_{[T\leqslant C]}$
siendo C la cota de censura, constante o variable para cada unidad, como explicaremos posteriormente y $I_{[A]}$ denota la función indicador del conjunto A, es decir:

$$X = \begin{cases} T & \text{si } T \leqslant C \\ C & \text{si } T \geqslant C \end{cases} \qquad \delta = \begin{cases} 0 & \text{si la observación es censurada} \\ 1 & \text{si la observación es no censurada} \end{cases}$$

Vagamente hablando, una observación censurada contiene únicamente una información parcial acerca de la variable aleatoria de interés.

Ahora bien, como hemos visto en los ejemplos anteriores las causas que originan la censura de una observación pueden -

ser aleatorias o controladas; ésto hace que distingamos entre tres tipos de censura siguiendo a Miller (1981):

<u>Censura de tipo I</u>

C es una constante para todas las unidades muestrales, preestablecida por el estadístico.

<u>Censura de tipo II</u>

La observación cesa después de un número $r < n$ predeterminado de muertes, de forma que C se convierte en una variable aleatoria (Ver Cox y Oakes; 1984).

Ambos tipos de censura surgen en numerosas aplicaciones de ingeniería.

<u>Censura aleatoria</u>

C es una variable aleatoria que se supone independiente de T. Este tipo de censura, denominada con más rigor censura aleatoria por la derecha, surge en las aplicaciones a la biología y la medicina.

Además este tipo de censura engloba como caso particular el tipo I.

Podría citarse, de forma similar, aunque con una aplicación más limitada, la censura aleatoria por la izquierda, cuando la información que se registra para algunas unidades es que su tiempo de vida es anterior a C, variable de censura independiente de T, con lo que la variable observada, en este caso, es: (X,δ) con $X = \text{máx}(T,C)$ y $\delta = I_{[C\leq T]}$.

Volviendo a nuestro objetivo inferencial, el estadístico dispone de un experimento censurado en el que observa la variable (X,δ). Pensemos que el estadístico puede elegir entre varias variables de censura, o lo que es lo mismo, entre varios experimentos censurados en los que observa X para diferentes variables de censura C. Un método natural para elegir el más apropiado sería comparar la información estadística contenida en dichos experimentos y elegir el de mayor información. Ahora bien, como es conocido, no hay una única medida de la información estadística contenida en un experimento.

Hollander, Proschan y Sconing (1985 a, 1985 b) consideran este problema para modelos censurados aleatoriamente por la derecha, en general no paramétricos, y sugieren que las medidas de información elegidas cumplan dos requisitos intuitivos:

i) Si $C_1 \overset{st}{\leqslant} C_2$ (*) (La variable de censura C_1 es estocásticamente más pequeña que la C_2) la información en (X_1,δ_1) es más pequeña que la información en (X_2,δ_2) para todo T, donde

$$X_i = \min(T,C_i) \quad y \quad \delta_i = I_{[T\leqslant C_i]} \quad i=1,2$$

ii) Para todo T y C la información en T es más grande que la información en (X,δ), $X = \min(T,C)$, $\delta = I_{[T\leqslant C]}$

(*) Dadas dos variables aleatorias Y,Z, $Y \overset{st}{\leqslant} Z \iff P(Y > y) \leqslant P(Z > y) \quad \forall y$ (Rohatgi; 1976, p. 554)

ii) se sigue de i) tomando $C_2=\infty$, de modo que X_2 tiene la misma distribución que T.

Los autores consideran que una medida de información que no verifique i) y ii) es inadecuada.

En (Hollander, Proschan y Sconing; 1985 a) los autores eligen la entropía de Shannon, cuando las variables T y C son discretas. Dado $p_i = P(T=i)$, $q_i = P(C=i)$, definen la información esperada para el experimento censurado discreto (X,δ) basado en (T,C) como:

$$H(p,q) = E_C\left[E_X(-\log P(X|C=i)\right]$$

ésto es, la media de las entropías de Shannon para los experimentos con variables de censura degeneradas en C=i. Es interesante observar que $H(p,q) = H(T)-H(T|X,\delta)$.

Muestran que esta medida verifica las condiciones i) y ii) anteriores.

Para el caso continuo, establecen que la entropía de Shannon no es satisfactoria pues en algunos casos no está definida y en otros puede ser negativa, y proponen una medida de la dispersión de las observaciones censuradas (X,δ), como medida de información que satisface las condiciones i) y ii). Para una variable no censurada, esta medida se reduce a la varianza de T.

Asímismo, en (Hollander, Proschan y Sconing; 1985 b) los autores proponen medidas de información adaptando medidas de dependencia entre las variables T y X, que verifican las condiciones i) y ii).

Goel (1987) proporciona un resultado más fuerte para este tipo de experimentos (censurados aleatoriamente por la derecha), para modelos paramétricos, que establece lo siguiente:

Dados dos experimentos censurados aleatoriamente por la derecha ε y ζ basados en las variables aleatorias (T,C_1) y (T,C_2) respectivamente, tales que $C_1 \overset{st}{\leq} C_2$, entonces el experimento ζ es suficiente para el experimento ε según la definición de Blackwell (1951, 1953).

Este resultado nos permite utilizar cualquier medida de -información que verifique la propiedad de suficiencia de experimentos, pues dicha propiedad nos conducirá a las condiciones i) y ii).

2.2. NOTACION Y ANTECEDENTES

Sea T la variable tiempo de vida en un estudio de supervivencia cuya distribución depende de un parámetro desconocido θ, $\theta \in \Theta$. Pensemos en la existencia de una variable de censura C de distribución conocida que impide la observación completa de T. Denotamos por X a la nueva variable observada, obtenida de T según C, cuya distribución dependerá de θ. Sean ε_o y ε_c los experimentos no censurado y censurado respectivamente que consisten en una observación de las variables respectivas T y X.

Denotaremos por $\varepsilon_o^{(n)}$ y $\varepsilon_c^{(n)}$ a los experimentos que consisten en la observación de n réplicas independientes de los experimentos ε_o y ε_c respectivamente. $I_o(\theta)$ e $I_c(\theta)$ denotarán una medida de información, acerca de θ, proporcionada por los experimentos ε_o y ε_c respectivamente. Análogamente $I_{n,o}(\theta)$ e $I_{n,c}(\theta)$ para los experimentos $\varepsilon_o^{(n)}$ y $\varepsilon_c^{(n)}$.

El fenómeno de la pérdida de información en modelos censurados ha sido estudiado entre otros por Brooks (1982), Barlow y Hsiung (1983), Hollander, Proschan y Sconing (1985a,1985b) y Goel (1987).

Nuestro interés se centra en evaluar esta pérdida de información, que se produce al observar ε_c en lugar de ε_o.

Brooks (1982) utilizando la medida de información de Shannon, sugiere dos caminos:

a) <u>La pérdida relativa de información</u> que se produce usando $\varepsilon_c^{(n)}$ en vez de $\varepsilon_o^{(n)}$ definida por

$$L_{n,c}=\frac{I_{n,o}(\theta)-I_{n,c}(\theta)}{I_{n,o}(\theta)} \tag{2.2.1}$$

b) <u>La eficiencia relativa</u> de $\varepsilon_c^{(n)}$ comparado con $\varepsilon_o^{(n)}$ para hacer inferencias sobre θ, definida por

$$R_{n,c}=\frac{n^*}{n}$$

donde n es el tamaño muestral de $\varepsilon_c^{(n)}$ y n^* es el tamaño muestral interpolado de $\varepsilon_o^{(n^*)}$ que se requiere para que $I_{n^*,o}(\theta)=$ $=I_{n,c}(\theta)$.

Este concepto, utilizado por Brooks (1980) para comparar dos tipos de experimentos con datos apareados, es análogo al de eficiencia relativa de un test no paramétrico comparado con su equivalente paramétrico, aquí se usan potencias de test en lugar de medidas de información (Hodges y Lehman; 1956).

Brooks (1982) utiliza ambos caminos en un contexto bayesiano, cuando la variable T se distribuye exponencialmente, la variable de censura C es de tipo I o tipo II y eligiendo como medida de información la medida de Shannon. Calcula asímismo la eficiencia relativa límite del experimento censurado respecto del no censurado, definida por $R_c=\lim_{n\to\infty} R_{n,c}$

Vamos a intentar ampliar estos conceptos, cuando θ es k--variante y la medida de información es la matriz de Fisher co rrespondiente. Supondremos un tipo de censura aleatoria por la derecha.

Sugerimos dos maneras de abordar este tema, que exponemos separadamente, la primera en las secciones 2.3 y 2.4,y la segunda en la sección 2.5.

2.3. MEDIDA MATRICIAL L_c

Sean $I_o(\theta)$ e $I_c(\theta)$ las matrices de información de Fisher, acerca de $\theta \in \Theta \subset \mathbb{R}^k$,obtenidas a partir de ε_o y ε_c respectivamente.

La idea será definir una medida matricial "natural" de la pérdida relativa de información y a partir de ésta, mediante - la elección de funciones reales adecuadas de dicha matriz, definir medidas reales de la pérdida relativa de información.

2.3.1. DEFINICION Y PROPIEDADES

Definición 2.3.1

La matriz

$$L_c(\theta) = (I_o(\theta)-I_c(\theta))I_o^{-1}(\theta) = I-I_c(\theta)I_o^{-1}(\theta)$$

siempre que esté definida, es una medida matricial de la pérdida relativa de información acerca de θ que se produce al observar ε_c en vez de ε_o.

Llamaremos a $L_c(\theta)$ matriz de pérdidas acerca de θ,al observar ε_c en vez de ε_o. Una vez fijada la distribución de la variable de censura, la matriz de pérdidas $L_c(\theta)$ es función de θ únicamente.

Es claro que la definición de $L_c(\theta)$ requiere que la matriz

$I_o(\theta)$ sea no singular para todo $\theta \in \Theta$, o lo que es equivalente (Propiedad A.2 apéndice), que $I_o(\theta)$ sea definida positiva ($I_o(\theta) > 0$) para todo $\theta \in \Theta$. Esta será pues la única limitación que condicionará todo el proceso que expondremos en esta sección.

A pesar del claro paralelismo de esta definición matricial con la correspondiente real de la pérdida relativa de información (Ver (2.2.1)), no hay un significado intuitivo sencillo de la matriz de pérdidas como medida de la pérdida de información, serán sus buenas propiedades y en particular las de sus autovalores las que irán dotando de contenido a dicha matriz $L_c(\theta)$. Expondremos en cuatro teoremas sus propiedades más importantes así como algunas consecuencias de las mismas en otros tantos corolarios.

La medida real definida en (2.2.1) es un número real comprendido entre 0 y 1, un resultado análogo en términos matriciales es el que expresa el siguiente teorema.

Teorema 1

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces:

i) La matriz $L_c(\theta)$ tiene todos sus autovalores comprendidos entre 0 y 1, para toda variable C y todo $\theta \in \Theta$

ii) $\lambda_i(L_c(\theta)) = 0,\ i=1,\dots,k \Leftrightarrow L_c(\theta) = \underset{\sim}{O}$ (matriz nula) $\Leftrightarrow$
$\Leftrightarrow I_o(\theta) = I_c(\theta)$

iii) $\lambda_i(L_c(\theta)) = 1\ ,\ i=1,\dots,k \Leftrightarrow L_c(\theta) = I$ (matriz unidad) $\Leftrightarrow$
$\Leftrightarrow I_c(\theta) = \underset{\sim}{O}$

<u>Demostración</u>

i) Por ser $I_o(\theta) > 0$ para todo $\theta \in \Theta$, entonces

$$I_o^{-1}(\theta) > 0 \quad \text{para todo} \quad \theta \in \Theta \qquad (2.3.1)$$

según el corolario A.7.2.2, p. 475 de (Mardia et al; 1979)
Por otra parte $\varepsilon_o \gtrsim \varepsilon_c$ para toda variable C (Goel;1987, cor. 3.2), por lo que

$$I_o(\theta)-I_c(\theta) \geqslant 0 \quad \text{para toda C y} \quad \theta \in \Theta \qquad (2.3.2)$$

por la propiedad (5) de la matriz de Fisher.
Por darse (2.3.1) y (2.3.2), aplicando el corolario A.7.3.1 p. 476 de (Mardia et al; 1979) resulta que todos los autovalores de $(I_o(\theta)-I_c(\theta))I_o^{-1}(\theta)$ son reales y no negativos, es decir

$$\lambda_i(L_c(\theta)) \geqslant 0 \quad i=1,\dots,k \qquad (2.3.3)$$

Por otra parte $I_c(\theta) \geqslant 0$ para toda C y $\theta \in \Theta$, ésto junto con (2.3.1) nos lleva utilizando el mismo argumento anterior a que:

todos los autovalores de $I_c(\theta)I_o^{-1}(\theta)$ son reales y no negativos, por lo que

$$\lambda_i(-I_c(\theta)I_o^{-1}(\theta)) \leqslant 0 \quad i=1,\dots,k$$

Los autovalores de la matriz de pérdidas $L_c(\theta)$ serán las soluciones λ de su ecuación característica

$$|L_c(\theta)-\lambda I| = 0 \iff |I-I_c(\theta)I_o^{-1}(\theta)-\lambda I| = 0 \iff$$

$$\iff |-I_c(\theta)I_o^{-1}(\theta)-(\lambda-1)I| = 0 \implies (\lambda-1) \leqslant 0$$

con lo que

$$\lambda_i(L_c(\theta)) \leq 1 \qquad i=1,\ldots,k \qquad (2.3.4)$$

Combinando las relaciones (2.3.3) y (2.3.4) se obtiene el resultado.

ii) Demostraremos que:

$$\lambda_i(L_c(\theta)) = 0 \quad , \quad i=1,\ldots,k \Leftrightarrow I_o(\theta) = I_c(\theta)$$

La última equivalencia de este apartado es inmediata de la definición 2.3.1.

$\Leftarrow$) Inmediata

$$\Rightarrow) \ \lambda_i(L_c(\theta)) = 0 \ , \quad i=1,\ldots,k \Leftrightarrow \lambda_i(I-I_c(\theta)I_o^{-1}(\theta))=0 \quad i=1,\ldots,k$$

$$\Leftrightarrow \lambda_i(I_c(\theta)I_o^{-1}(\theta)) = 1 \qquad i=1,\ldots,k$$

entonces

$$1 = \prod_{i=1}^{k} \lambda_i(I_c(\theta)I_o^{-1}(\theta)) = |I_c(\theta)I_o^{-1}(\theta)| =$$

$$= |I_c(\theta)||I_o^{-1}(\theta)| = |I_c(\theta)||I_o(\theta)|^{-1}$$

es decir

$|I_c(\theta)| = |I_o(\theta)| \neq 0$ lo que equivale a que

$$0 \neq \prod_{i=1}^{k} \lambda_i(I_c(\theta)) = \prod_{i=1}^{k} \lambda_i(I_o(\theta)) \qquad (2.3.5)$$

Por otra parte

$$I_o(\theta)-I_c(\theta) \geq 0$$

de donde

$$\lambda_i(I_c(\theta)) \leqslant \lambda_i(I_o(\theta)) \qquad i=1,\ldots,k \qquad (2.3.6)$$

(Bellman; 1970, p. 117)

De (2.3.5) y (2.3.6) resulta que

$$\lambda_i(I_c(\theta)) = \lambda_i(I_o(\theta)) \qquad i=1,\ldots,k$$

entonces

$$\text{Traza}(I_o(\theta)-I_c(\theta)) = \text{Traza}(I_o(\theta))-\text{Traza}(I_c(\theta)) = 0$$

por lo que

$$\lambda_i(I_o(\theta)-I_c(\theta)) = 0 \qquad i=1,\ldots,k \qquad (2.3.7)$$

Por ser $I_o(\theta)-I_c(\theta)$ simétrica, en virtud de su descomposición espectral (2.3.7) es equivalente a

$$I_o(\theta)-I_c(\theta) = \underset{\sim}{0} \iff I_o(\theta) = I_c(\theta)$$

iii) Demostraremos que:

$$\lambda_i(L_c(\theta)) = 1 \ , \quad i=1,\ldots,k \iff I_c(\theta) = \underset{\sim}{0}$$

La última equivalencia de este apartado es inmediata de la definición 2.3.1.

$\Leftarrow$) Inmediata

$\Rightarrow$) $\lambda_i(L_c(\theta)=1 \quad i=1,\ldots,k \iff \lambda_i(I_c(\theta)I_o^{-1}(\theta)) = 0 \quad i=1,\ldots,k$

Sabemos que

$$I_o^{-1}(\theta) > 0 \quad \text{e} \quad I_c(\theta) \geqslant 0 \quad \forall\ \theta \in \Theta$$

entonces el corolario 2.2.1 de (Anderson y Das Gupta; 1963) implica que

$$\lambda_k(I_o^{-1}(\theta))\lambda_i(I_c(\hat{\theta})) \leq \lambda_i(I_c(\hat{\theta})I_o^{-1}(\theta)) = 0 \qquad i=1,\ldots,k$$

por lo que

$$\lambda_i(I_c(\hat{\theta})) = 0 \qquad i=1,\ldots,k$$

y como $I_c(\theta)$ es simétrica, en virtud de su descomposición espectral resulta que

$$I_c(\hat{\theta}) = \underset{\sim}{0}$$

□

En numerosas ocasiones la matriz de Fisher es diagonal, este hecho es el que inspira los dos siguientes corolarios. En ambos suponemos implícitamente que $I_o(\hat{\theta}) > 0$.

Corolario 1

Si $L_c(\theta)$ es una matriz simétrica, entonces

$L_c(\hat{\theta})$ es semidefinida positiva.

Demostración

Una condición necesaria y suficiente para que una matriz simétrica sea semidefinida positiva es que todos sus autovalores sean no negativos (Bellman; 1970, p. 54, th. 3).

El teorema 1i) completa la demostración.

□

Corolario 2

Si $I_c(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$, entonces

La matriz $L_c(\theta)$ es diagonal, semidefinida positiva con elemento i-ésimo la pérdida relativa de información acerca de θ_i (componente i-ésima de θ) para todo $\theta \in \Theta$.

Demostración

Siguiendo la notación del capítulo I, $I_o(\theta_i)$ expresa la medida de información de Fisher (unidimensional) acerca del parámetro θ_i, a partir de ε_o, dados los valores del resto de parámetros perturbadores θ_j, $j \neq i$. Análogamente para $I_c(\theta_i)$.

La matriz $I_o(\theta)$ es la particularización de $I_c(\theta)$ cuando la variable de censura $C=+\infty$. Entonces $I_o(\theta)$ es diagonal con elementos diagonales $I_o(\theta_i)$ $i=1,\dots,k$. Además $I_o(\theta_i) > 0$ $i=1,\dots,k$, por ser $I_o(\theta) > 0$.

Entonces la matriz $I_o^{-1}(\theta) > 0$ es una matriz diagonal, siendo sus elementos diagonales $\frac{1}{I_o(\theta_i)} > 0$ $i=1,\dots,k$.

Por otra parte sabemos que

$$I_o(\theta)-I_c(\theta) \geq 0 \quad \text{por ser } \varepsilon_o \gtrsim \varepsilon_c \quad \text{(Goel;1987,corolario 3.}$$

y ésto es equivalente a decir que

$$I_o(\theta_i) \geq I_c(\theta_i) \quad i=1,\dots,k$$

por consiguiente $I_c(\theta)I_o^{-1}(\theta)$ es una matriz diagonal y su elemento i-ésimo es $\frac{I_c(\theta_i)}{I_o(\theta_i)}$ que verifica $0 \leq \frac{I_c(\theta_i)}{I_o(\theta_i)} \leq 1$

De esta forma, la matriz $L_c(\theta) = I-I_c(\theta)I_o^{-1}(\theta)$ es una matriz diagonal, semidefinida positiva con elemento i-ésimo

$$1 - \frac{I_c(\theta_i)}{I_o(\theta_i)} = \frac{I_o(\theta_i)-I_c(\theta_i)}{I_o(\theta_i)}$$

$\square$

Nuestro siguiente objetivo será comparar las matrices de pérdidas originadas por dos o más variables de censura ordenadas estocásticamente.

Cuando utilizamos una medida de información real, el resultado deseado sería:

Dadas C_1 y C_2 dos variables de censura tales que $C_1 \overset{st}{\leqslant} C_2$ entonces

$$L_{C_1} \geqslant L_{C_2}$$

donde L_{C_i} denota la pérdida relativa de información (Ver (2.2.1)) al observar ε_{C_i} en vez de ε_o, siendo ε_{C_i} el experimento censurado según la variable C_i, i=1,2. Un resultado equivalente en términos matriciales podría ser el siguiente teorema.

Teorema 2

Sean $C_1 \overset{st}{\leqslant} C_2$ dos variables de censura e $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces:

i) La matriz $L_{C_1}(\theta)-L_{C_2}(\theta)$ tiene todos sus autovalores no negativos para todo $\theta \in \Theta$

ii) $\lambda_i(L_{C_1}(\theta)) \geqslant \lambda_i(L_{C_2}(\theta))$ i=1,...,k para todo $\theta \in \Theta$

Demostración

i) Por ser $C_2 \overset{st}{\geqslant} C_1$ se cumple

$$\varepsilon_{C_2} \geq \varepsilon_{C_1} \quad \text{(Goel; 1987, th. 3.1)}$$

entonces

$$I_{c_2}(\theta) \geq I_{c_1}(\theta) \quad \forall\theta \in \Theta$$ **(propiedad (5) de la matriz de Fisher)**

es decir

$$I_{c_2}(\theta)-I_{c_1}(\theta) \geq 0 \qquad (2.3.8)$$

por tanto

$$L_{c_1}(\theta)-L_{c_2}(\theta) = I_{c_2}(\theta)I_o^{-1}(\theta)-I_{c_1}(\theta)I_o^{-1}(\theta) =$$

$$= (I_{c_2}(\theta)-I_{c_1}(\theta))I_o^{-1}(\theta)$$

Ya hemos visto anteriormente que $I_o^{-1}(\theta) > 0$ si $I_o(\theta) > 0$, este hecho y (2.3.8) proporcionan el resultado sin más que aplicar el corolario A.7.3.1, p. 476 de (Mardia et al; 1979).

ii) $\lambda_i(L_{c_1}(\theta)) = \lambda_i(I-I_{c_1}(\theta)I_o^{-1}(\theta)) \qquad i=1,\ldots,k$

Puesto que

$$|I-I_{c_1}(\theta)I_o^{-1}(\theta)-\lambda I| = 0 \iff |I_{c_1}(\theta)I_o^{-1}(\theta)+(\lambda-1)I| = 0$$

entonces

$$\lambda_i(L_{c_1}(\theta)) = 1-\lambda_i(I_{c_1}(\theta)I_o^{-1}(\theta)) \qquad i=1,\ldots,k \qquad (2.3.9)$$

Por otra parte, teniendo en cuenta que $I_o^{-1}(\theta)=I_o^{-1/2}(\theta)I_o^{-1/2}(\theta)$

$$\lambda_i(I_{c_1}(\theta)I_o^{-1}(\theta)) = \lambda_i(I_o^{-1/2}(\theta)I_{c_1}(\theta)I_o^{-1/2}(\theta))$$

$$i=1,\ldots,k \qquad (2.3.10)$$

según el teorema A.6.2 p. 468 de (Mardia et al; 1979).

Análogamente

$$\lambda_i(L_{c_2}(\theta)) = 1-\lambda_i(I_{c_2}(\theta)I_o^{-1}(\theta)) \quad i=1,\ldots,k \qquad (2.3.11)$$

y $\lambda_i(I_{c_2}(\theta)I_o^{-1}(\theta)) = \lambda_i(I_o^{-1/2}(\theta)I_{c_2}(\theta)I_o^{-1/2}(\theta)) \quad i=1,\ldots,k$

(2.3.12)

Además, por hipótesis $I_{c_2}(\theta) \geq I_{c_1}(\theta)$ (Ver (2.3.8))

Entonces, según el ejercicio 12(d) p. 93 de (Bellman;1970)

$$I_o^{-1/2}(\theta)I_{c_2}(\theta)I_o^{-1/2}(\theta) \geq I_o^{-1/2}(\theta)I_{c_1}(\theta)I_o^{-1/2}(\theta)$$

Este resultado, las relaciones (2.3.10), (2.3.12) y el teorema 3 p. 117 de (Bellman; 1970) implican que

$$\lambda_i(I_{c_2}(\theta)I_o^{-1}(\theta)) \geq \lambda_i(I_{c_1}(\theta)I_o^{-1}(\theta)) \qquad i=1,\ldots,k$$

A partir de aquí, las condiciones (2.3.9) y (2.3.11) conducen de manera inmediata al resultado.

$\square$

A continuación estudiaremos los efectos que una transformación biyectiva del parámetro θ tiene en la matriz de pérdidas $L_c(\theta)$.

Cuando utilizamos una medida de información real, una buena propiedad de la medida real L_c sería la siguiente:

Si $\phi = f(\theta)$ es una transformación biyectiva de θ, entonces $L_c(\theta) = L_c(\phi)$ para toda variable de censura C, es decir "la pérdida relativa de información es invariante bajo transformaciones biyectivas del parámetro".

El teorema 3 pretende dar un resultado equivalente en términos matriciales.

Sea $\phi = f(\theta)$ una transformación biyectiva de θ y $A = (a_{ij})$ con $a_{ij} = \dfrac{\partial\theta_i}{\partial\phi_j}$ $i,j=1,\ldots,k$ denotaremos por $I_c(\phi)$, $I_o(\phi)$ y $L_c(\phi)$ a las matrices correspondientes respecto del parámetro ϕ.

Teorema 3

Si $I_o(\theta) > 0$ y A es no singular para todo $\theta \in \Theta$, entonces

Las matrices $L_c(\phi)$ y $L_c(\theta)$ tienen la misma ecuación característica para toda C y $\theta \in \Theta$ y por tanto los mismos autovalores.

Demostración

Sabemos que $I_c(\phi) = A'I_c(\theta)A$ para toda C (Ver (1.3.8)
y en particular $I_o(\phi) = A'I_o(\theta)A$

Por hipótesis $I_o(\theta)$ y A son no singulares, por lo que $I_o(\phi)$ es no singular

Por definición $L_c(\phi) = I-I_c(\phi)I_o^{-1}(\phi)$ entonces

$$L_c(\phi) = I-A'I_c(\theta)AA^{-1}I_o^{-1}(\theta)A'^{-1} = I-A'I_c(\theta)I_o^{-1}(\theta)A'^{-1}$$

llamando B=A' podemos escribir

$$L_c(\phi) = B(I-I_c(\theta)I_o^{-1}(\theta))B^{-1} = BL_c(\theta)B^{-1}$$ por lo que

las matrices $L_c(\phi)$ y $L_c(\theta)$ son similares (Rao; 1973, p. 76) y por lo tanto, tienen la misma ecuación característica.

□

Corolario 3

Si $I_c(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$, entonces

Los autovalores de la matriz $L_c(\phi)$ son las pérdidas relativas de información acerca de cada componente θ_i de θ.

Demostración

Por el corolario 2 la matriz $L_c(\theta)$ es diagonal por lo que sus autovalores son sus elementos diagonales. El teorema 3 completa la demostración.

□

Como consecuencia de este corolario:

Siempre que podamos transformar el parámetro ϕ mediante - una transformación biyectiva en otro θ que haga diagonal la matriz de información de Fisher, la matriz de pérdidas (respecto de θ o ϕ) debida a la censura tiene por autovalores las pérdidas relativas de información acerca de cada componente de θ.

Corolario 4

Sean $C_1 \overset{st}{\leq} C_2$ dos variables de censura, $I_o(\theta) > 0$ y A no singular,para todo $\theta \in \Theta$, entonces

Las matrices $L_{c_1}(\theta)-L_{c_2}(\theta)$ y $L_{c_1}(\phi)-L_{c_2}(\phi)$ tienen los mismos autovalores.

Demostración

Por el teorema 3

$$L_{c_i}(\phi) = BL_{c_i}(\theta)B^{-1} \qquad i=1,2$$

entonces

$$L_{c_1}(\phi)-L_{c_2}(\phi) = B(L_{c_1}(\theta)-L_{c_2}(\theta))B^{-1}$$

siguiendo el mismo argumento del teorema 3 llegamos al resultado.

□

Seguidamente, estudiamos el comportamiento de la matriz de pérdidas cuando los experimentos observados son $\varepsilon_o^{(n)}$ y $\varepsilon_c^{(n)}$.

Denotamos por $L_{n,c}(\theta)$ a la matriz de pérdidas acerca de θ al observar $\varepsilon_c^{(n)}$ en vez de $\varepsilon_o^{(n)}$, es decir:

$$L_{n,c}(\theta) = I-I_{n,c}(\theta)I_{n,o}^{-1}(\theta)$$

Teorema 4

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$, entonces

$L_{n,c}(\theta) = L_c(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Demostración

$I_{n,c}(\theta) = nI_c(\theta)$ por la propiedad (2) de aditividad de la matriz de Fisher. Igualmente $I_{n,o}(\theta) = nI_o(\theta)$ es definida positiva por serlo $I_o(\theta)$

entonces

$$I_{n,o}^{-1}(\theta) = n^{-1}I_o^{-1}(\theta) \quad \text{de donde}$$

$$L_{n,c}(\theta) = I-(nI_c(\theta))(n^{-1}I_o^{-1}(\theta)) = I-I_c(\theta)I_o^{-1}(\theta) = L_c(\theta)$$

$\square$

Los teoremas 1,2,3 y 4 tienen su particularización, cuando θ es univariante, en los cuatro próximos teoremas.

Sea $\theta \in \Theta \subset \mathbb{R}$, $I_o^*(\theta)$, $I_c^*(\theta)$ la medida de información de Fisher (unidimensional) acerca de θ, obtenida a partir de ε_o y ε_c respectivamente.

Sea $L_c^*(\theta) = \dfrac{I_o^*(\theta)-I_c^*(\theta)}{I_o^*(\theta)}$ el número real que mide la pérdida relativa de información producida al observar ε_c en vez de ε_o, para cada $\theta \in \Theta$.

Análogamente $I_{n,o}^*(\theta)$, $I_{n,c}^*(\theta)$ y $L_{n,c}^*(\theta)$ para los experimentos $\varepsilon_o^{(n)}$ y $\varepsilon_c^{(n)}$.

Sean $C_1 \overset{st}{\leq} C_2$ dos variables de censura ordenadas estocásticamente y $\phi=f(\theta)$ una transformación biyectiva de θ.

Suponemos $I_o^*(\theta) > 0$ para todo $\theta \in \Theta$

Teorema * 1

i) $0 \leqslant L_c^*(\theta) \leqslant 1$ para toda variable C y $\theta \in \Theta$

ii) $0 = L_c^*(\theta) \Leftrightarrow I_o^*(\theta) = I_c^*(\theta)$

iii) $1 = L_c^*(\theta) \Leftrightarrow I_c^*(\theta) = 0$

Teorema * 2

$L_{c_1}^*(\theta) \geqslant L_{c_2}^*(\theta)$ para todo $\theta \in \Theta$

Teorema * 3

$L_c^*(\theta) = L_c^*(\phi)$ para toda variable C

Teorema * 4

$L_{n,c}^*(\theta) = L_c^*(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Demostraciones

Se apoyan en los siguientes resultados

$\varepsilon_o \geq \varepsilon_c$ para cualquier variable C de censura aleatoria por la derecha (Goel; 1987, cor. 3.2)

$\varepsilon_X \geq \varepsilon_Y \Rightarrow I_X^*(\theta) \geqslant I_Y^*(\theta)$ (Stone; 1961, th.1)

$\varepsilon_{c_2} \geq \varepsilon_{c_1}$ (Goel; 1987, th. 3.1)

$I_X^*(\phi) = (\frac{\partial\theta}{\partial\phi})^2 I_X^*(\theta)$ (Fisher; 1956, p. 155)

$I_{n,X}^*(\theta) = nI_X^*(\theta)$ Propiedad de aditividad (Fisher; 1956, p. 149)

$\square$

2.3.2. PERDIDA RELATIVA DE INFORMACION. MEDIDAS REALES

Tras este paréntesis para el caso de θ univariante, volvemos al caso que nos ocupa, es decir $\theta \in \Theta \subset \mathbb{R}^k$ y matriz de Fisher como medida de información. Vamos a definir tres medidas reales de la pérdida relativa de información a partir de tres funciones reales de la matriz de pérdidas. Exigiremos a estas medidas que tengan buenas propiedades, en concreto, las que acabamos de enumerar en los teoremas * 1 a 4, a las que nos referiremos como propiedades 1 a 4.

Cualquier función real de la matriz de pérdidas verificará la propiedad 4, al coincidir las matrices $L_{n,c}(\theta)$ y $L_c(\theta)$ (teorema 4).

La propiedad 3 se cumple para toda medida real que se defina como función de los autovalores, pues las matrices $L_c(\theta)$ y $L_c(\hat{\theta})$ tienen los mismos autovalores (teorema 3). Debemos pues elegir entre éstas, las que cumplan las propiedades 1 y 2.

<u>Definición 2.3.2</u>

La función real

$$L_c^1(\theta) = \frac{1}{k} \sum_{i=1}^{k} \lambda_i(L_c(\theta))$$

siempre que esté definida, es una medida real de la pérdida relativa de información que, acerca de θ, se produce al observar ε_c en vez de ε_o.

La medida $L_c^1(\theta)$ es la media aritmética de los autovalores de la matriz de pérdidas $L_c(\theta)$, y está definida siempre que lo esté dicha matriz, es decir cuando $I_o(\theta) > 0$ para todo $\theta \in \Theta$.

Una expresión alternativa de $L_c^1(\theta)$ será:

$$L_c^1(\theta) = \frac{1}{k} \text{ Traza}[L_c(\theta)]$$

Denotaremos por $L_{n,c}^1(\theta)$ a la generalización de la medida anterior para los experimentos $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$, es decir

$$L_{n,c}^1 = \frac{1}{k} \text{ Traza}[L_{n,c}(\theta)]$$

Debemos probar que la medida real $L_c^1(\theta)$ verifica las propiedades 1 a 4; el próximo teorema recoge este resultado.

Sean $C_1 \overset{st}{\leqslant} C_2$ dos variables de censura y $\phi = f(\theta)$ una transformación biyectiva de θ con $A = (\frac{\partial\theta_i}{\partial\phi_j})$ $i,j=1,\ldots,k$

Teorema 5

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

a) $\begin{cases} \text{i)} & 0 \leqslant L_c^1(\theta) \leqslant 1 \quad \text{para toda variable } C \text{ y } \theta \in \Theta \\ \text{ii)} & 0 = L_c^1(\theta) \Leftrightarrow I_o(\theta) = I_c(\theta) \\ \text{iii)} & 1 = L_c^1(\theta) \Leftrightarrow I_c(\theta) = \underset{\sim}{0} \end{cases}$

b) $L_{C_1}^1(\theta) \geqslant L_{C_2}^1(\theta)$ para todo $\theta \in \Theta$

c) $L_{n,c}^1(\theta) = L_c^1(\theta)$ para todo $n \in N$, toda variable C y $\theta \in \Theta$

Si además, la matriz A es no singular para todo $\theta \in \Theta$

d) $L_c^1(\theta) = L_c^1(\phi)$ para toda variable C.

Demostración

La definición 2.3.2 de $L_c^1(\theta)$ y los teoremas 1,2 ii), 4 y 3 conducen a los resultados a) b) c) y d) respectivamente.

□

En el caso particular de que $I_c(\theta)$ sea diagonal para toda variable C, tenemos

Corolario 5

Si $I_c(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$, entonces $L_c^1(\theta)$ es la media aritmética de las pérdidas relativas de información acerca de cada θ_i (componente i-ésima de θ) para todo $\theta \in \Theta$

Demostración

Es consecuencia inmediata de la definición 2.3.2 y del Corolario 2.

□

Definimos, a continuación, una segunda medida real de la pérdida relativa de información, basada como la anterior en los autovalores de la matriz de pérdidas.

Definición 2.3.3

La función real

$$L_c^2(\theta) = \left[\prod_{i=1}^{k} \left[1+\lambda_i(L_c(\theta))\right]\right]^{1/k} -1$$

siempre que esté definida, es una medida real de la pérdida relativa de información que, acerca de θ, se produce al observar ε_c en vez de ε_o.

La medida $L_c^2(\theta)$ es la media geométrica de los autovalores de la matriz $I+L_c(\theta)$ menos uno o también la φ-media de los autovalores de la matriz $L_c(\theta)$, correspondiente a la función $\varphi(x)=\log(1+x)$ (Calot; 1970, p. 73). Al igual que $L_c^1(\theta)$, la definición de $L_c^2(\theta)$ requiere que $I_o(\theta) > 0$ para todo $\theta \in \Theta$.

Una expresión alternativa de $L_c^2(\theta)$ será:

$$L_c^2(\theta) = |I+L_c(\theta)|^{1/k}-1$$

Como es habitual, denotaremos por $L_{n,c}^2(\theta)$ a la medida anterior relativa a los experimentos $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$, es decir

$$L_{n,c}^2(\theta) = |I+L_{n,c}(\theta)|^{1/k} -1$$

Veamos que la medida $L_c^2(\theta)$ verifica las propiedades 1 a 4.

Sean $C_1 \overset{st}{\leqslant} C_2$ dos variables de censura y $\phi=f(\theta)$ una transformación biyectiva de θ con $A = \left(\dfrac{\partial\theta_i}{\partial\phi_j}\right)$ $i,j=1,\ldots,k$.

Teorema 6

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

$$\text{a)} \begin{cases} \text{i)} & 0 \leqslant L_c^2(\theta) \leqslant 1 \quad \text{para toda variable } C \text{ y } \theta \in \Theta \\ \text{ii)} & 0 = L_c^2(\theta) \iff I_o(\theta) = I_c(\theta) \\ \text{iii)} & 1 = L_c^2(\theta) \iff I_c(\theta) = \underset{\sim}{0} \end{cases}$$

b) $L_{c_1}^2(\theta) \geqslant L_{c_2}^2(\theta)$ para todo $\theta \in \Theta$

c) $L_{n,c}^2(\theta) = L_c^2(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Si además, la matriz A es no singular para todo $\theta \in \Theta$

d) $L_c^2(\theta) = L_c^2(\phi)$ para toda variable C.

Demostración

La definición 2.3.3 de $L_c^2(\theta)$ y los teoremas 1, 2 ii), 4 y 3 conducen a los resultados a) b) c) y d) respectivamente.

$\square$

Daremos, por último, una tercera medida real de la pérdida relativa de información, basada, como las anteriores, en los autovalores de la matriz de pérdidas.

<u>Definición 2.3.4</u>

La función real

$$L_c^3(\theta) = \left[\frac{\sum_{i=1}^{k} [\lambda_i(L_c(\theta))]^2}{k} \right]^{1/2}$$

siempre que esté definida, es una medida real de la pérdida relativa de información que, acerca de θ, se produce al observar ε_c en vez de ε_o.

La medida $L_c^3(\theta)$ es por definición la media cuadrática de los autovalores de la matriz de pérdidas $L_c(\theta)$ y, como es habitual, requiere para su definición que la matriz $I_o(\theta) > 0$ para todo $\theta \in \Theta$.

El corolario 2 nos mostraba que si la matriz de Fisher $I_c(\theta)$ es diagonal, la matriz de pérdidas $L_c(\theta)$ también lo es; en este caso y en general si la matriz de pérdidas $L_c(\theta)$ es simétrica, la medida real $L_c^3(\theta)$ puede expresarse alternativamente como la norma euclídea de la matriz $\frac{1}{\sqrt{k}} L_c(\theta)$. Veámoslo

$$L_c^3(\theta) = \frac{1}{\sqrt{k}} \left[\sum_{i=1}^{k} [\lambda_i(L_c(\theta))]^2 \right]^{1/2} = \frac{1}{\sqrt{k}} \, ||L_c(\theta)|| =$$

$$= ||\frac{1}{\sqrt{k}} L_c(\theta)||$$

Es sencillo comprobar que esta medida verifica las propiedades 1 a 4.

Terminaremos esta sección 2.3 recogiendo algunas observaciones referentes a la matriz de pérdidas L_c y a sus medidas reales asociadas L_c^i $i=1,2,3.$

Observaciones

1.- La definición 2.3.1 establecía la expresión de la matriz de pérdidas L_c mediante un producto de dos matrices. No hay razón alguna que justifique el orden en que multiplicamos ambas matrices. Si conmutamos dicho producto obtenemos otra medida matricial estrechamente relacionada con L_c. Claramente

$$I_o^{-1}(I_o-I_c) = L_c'$$

Sólo en el caso de que la matriz L_c es simétrica, ambas medidas matriciales coinciden. En generàl pues, deberíamos elegir entre L_c y L_c' como matriz de pérdidas. Esta aparente ambigüedad, sin embargo no es tal, ya que tanto las propiedades como las definiciones que hemos visto se basan en los autovalores de L_c, y sabemos que

$$\lambda_i(L_c) = \lambda_i(L_c') \qquad i=1,\ldots,k$$

2.- Por definición, las medidas reales L_c^i $i=1,2,3$ son tres promedios diferentes de los autovalores de la matriz de pérdidas L_c, por lo que podemos escribir que

$$\lambda_k(L_c) \leq L_c^i \leq \lambda_1(L_c) \quad \text{para toda } C, \; i=1,2,3$$

Asímismo, debido a la relación entre las medias geométrica, aritmética y cuadrática (Calot; 1970, pp. 73-77) se verifica que

$$L_C^2(\theta) \leqslant L_C^1(\theta) \leqslant L_C^3(\theta) \quad \text{para toda } C \text{ y } \theta \in \Theta$$

dándose las dos igualdades a la vez, si y sólo si, para θ y C fijos, todos los autovalores de la matriz de pérdidas coinciden.

La discusión, en cuanto a la elección de una de estas tres medidas, se reduce pues, en cada caso, al análisis de la representatividad, ventajas e inconvenientes de estos promedios.

3.- Si la matriz de Fisher $I_C(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$, entonces la matriz de pérdidas L_C también es diagonal y las medidas reales L_C^i i=1,2,3 son los promedios correspondientes de las pérdidas relativas de información acerca de cada θ_i (componente i-ésima de θ) para todo $\theta \in \Theta$; L_C^1 la media aritmética, L_C^2 la φ- media, con $\varphi(x) = \log(1+x)$ y L_C^3 la media cuadrática.

2.4. MEDIDA MATRICIAL R_c

2.4.1. INTRODUCCION

En esta sección vamos a introducir una medida matricial "natural" de la eficiencia relativa de un experimento censurado $\varepsilon_c^{(n)}$ respecto al correspondiente no censurado $\varepsilon_o^{(n)}$, cuando la medida de información utilizada es la matriz de información de Fisher.

Dicha medida pretende ser la generalización matricial de la medida definida por Brooks (1982) cuando la medida de información elegida es la medida de Shannon. Recordamos esta definición:

La eficiencia relativa de $\varepsilon_c^{(n)}$ comparado con $\varepsilon_o^{(n)}$, para - hacer inferencias acerca de θ, viene definida, una vez fijada la variable de censura, por la siguiente función de n

$$R_{n,c} = \frac{n^*}{n}$$

siendo n el tamaño muestral de $\varepsilon_c^{(n)}$ y n^* el tamaño muestral interpolado de $\varepsilon_o^{(n^*)}$ que se requiere para que $I_{n^*,o}(\theta) = I_{n,c}(\theta)$.

Ahora bien, si la medida de información elegida fuese una medida paramétrica real y como tal dependiese de $\theta \in \Theta \subset \mathbb{R}$, podríamos extender la definición anterior de manera obvia como sigue

$R_{n,c}(\theta) = \frac{n^*(\theta)}{n}$ que, para n y C fijos, es una función de θ. Si además la medida de información elegida fuese aditiva para observaciones independientes, resulta que

$I_{n,c}(\theta) = nI_c(\theta)$ para toda variable de censura C y todo $\theta \in \Theta$ con lo que

$$I_{n^*,o}(\theta) = I_{n,c}(\theta) \Leftrightarrow \frac{n^*(\theta)}{n} = \frac{I_c(\theta)}{I_o(\theta)}$$

y por tanto

$$R_{n,c}(\theta) = \frac{I_c(\theta)}{I_o(\theta)} = R_c(\theta) \qquad \forall n \in \mathbb{N} \qquad (2.4.1)$$

es decir, para θ fijo, la eficiencia relativa de $\varepsilon_c^{(n)}$ comparado con $\varepsilon_o^{(n)}$ es constante para todo n y en particular, igual a la eficiencia relativa de ε_c comparado con ε_o, que denotamos por $R_c(\theta)$.

Además, existe una relación muy estrecha entre esta medida $R_c(\theta)$ de la eficiencia relativa y la correspondiente de la pérdida relativa de información $L_c(\theta)$. Es claro que

$$R_c(\theta) = 1-L_c(\theta) \quad \text{para toda variable de censura C y todo } \theta \in \Theta \qquad (2.4.2)$$

Supongamos que la medida de información aditiva elegida es la medida de Fisher (unidimensional). Denotemos por $I_o^*(\theta)$ e $I_c^*(\theta)$ a dicha medida de información, acerca de θ, para los experimentos ε_o y ε_c respectivamente.

Sea $R_c^*(\theta) = \dfrac{I_c^*(\theta)}{I_o^*(\theta)}$ la medida real de la eficiencia relativa, acerca de θ, de ε_c respecto de ε_o . Esta medida verifica las siguientes cuatro propiedades:

<u>Propiedad * 1</u>

i) $0 \leq R_c^*(\theta) \leq 1$ para toda variable C, y $\theta \in \Theta$

ii) $1 = R_c^*(\theta) \Leftrightarrow I_o^*(\theta) = I_c^*(\theta)$

iii) $0 = R_c^*(\theta) \Leftrightarrow I_c^*(\theta) = 0$

Propiedad * 2

Si $C_1 \overset{st}{\leqslant} C_2$ entonces

$$R^*_{C_2}(\theta) \geqslant R^*_{C_1}(\theta) \quad \text{para todo} \quad \theta \in \Theta$$

Propiedad * 3

Si $\phi = f(\theta)$ es una transformación biyectiva de θ, entonces

$$R^*_C(\theta) = R^*_C(\phi) \quad \text{para toda variable C}$$

Propiedad * 4

$$R^*_{n,C}(\theta) = R^*_C(\theta) \quad \text{para todo } n \in \mathbb{N}, \text{ toda variable C y } \theta \in \Theta$$

Demostraciones

La expresión (2.4.2) particularizada para la medida de información de Fisher $I^*_C(\theta)$ y los teoremas * 1, 2 y 3 conducen de manera inmediata a las propiedades * 1, 2 y 3. La propiedad * 4 es la particularización de la expresión (2.4.1) para la medida $I^*_C(\theta)$.

$\square$

Veamos ahora qué ocurre cuando $\theta \in \Theta \subset \mathbb{R}^k$ y la medida de información es la matriz de Fisher correspondiente. A partir de aquí, $I_o(\theta)$ e $I_c(\theta)$ denotarán, de nuevo, las matrices de información de Fisher, acerca de θ, para los experimentos ε_o y ε_c - respectivamente y análogamente $I_{n,o}(\theta)$ e $I_{n,c}(\theta)$ para $\varepsilon_o^{(n)}$ y $\varepsilon_c^{(n)}$ respectivamente. La matriz de Fisher posee la propiedad de aditividad para observaciones independientes, por lo que

$$I_{n,c}(\theta) = nI_c(\theta) \quad \text{para toda C y} \quad \theta \in \Theta$$

Nuestro primer paso, por analogía con el caso univariante, debería ser encontrar, para n fijo, el valor $n^*(\theta)$ que verifica la igualdad matricial siguiente

$$n^*(\theta)I_o(\theta) = nI_c(\theta) \Leftrightarrow \frac{n^*(\theta)}{n} I_o(\theta) = I_c(\theta)$$

Es claro que la ecuación anterior no tiene solución salvo que todos los elementos de la matriz $I_c(\theta)$ se obtengan de los correspondientes de $I_o(\theta)$ multiplicados por $\mu(\theta) \in \mathbb{R}$.

En analogía con el proceso seguido con la pérdida relativa de información (sección 2.3) definiremos primeramente <u>una medida matricial</u> de la eficiencia relativa para,posteriormente y basadas en esta matriz, definir <u>medidas reales</u> de la eficiencia relativa.

2.4.2 DEFINICION Y PROPIEDADES

<u>Definición 2.4.1</u>

La matriz

$$R_c(\theta) = I-L_c(\theta) = I_c(\theta)I_o^{-1}(\theta)$$

siempre que esté definida, es una medida matricial de la eficiencia relativa del experimento ε_c comparado con ε_o, para hacer inferencias acerca de θ.

Llamaremos a $R_c(\theta)$ matriz de eficiencias, acerca de θ, del experimento ε_c comparado con ε_o.

Es obvio por la definición anterior que la matriz $R_c(\theta)$ está definida siempre que lo esté $L_c(\theta)$ y por tanto siempre que la matriz <u>$I_o(\theta)$ sea no singular</u> para todo $\theta \in \Theta$.

Una vez fijada la variable de censura C, la matriz de eficiencias $R_c(\theta)$ depende únicamente de θ.

Las propiedades de la matriz de eficiencias serán fiel reflejo de las de la matriz de pérdidas. Los corolarios 6 a 9 recogen sus cuatro propiedades más importantes. Estas propiedades vienen a ser la generalización matricial de las correspondientes al caso univariante (propiedades * 1 a 4). Así, la propiedad * 1 tiene su equivalente matricial en el siguiente corolario.

Corolario 6

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

i) La matriz $R_c(\theta)$ tiene todos sus autovalores comprendidos entre 0 y 1, para toda variable C y todo $\theta \in \Theta$

ii) $\lambda_i(R_c(\theta)) = 1$, $i=1,\dots,k \iff R_c(\theta) = I \iff I_o(\theta) = I_c(\theta)$

iii) $\lambda_i(R_c(\theta)) = 0$, $i=1,\dots,k \iff R_c(\theta) = \underset{\sim}{0} \iff I_c(\theta) = \underset{\sim}{0}$

Demostración

Por definición, $R_c(\theta) = I-L_c(\theta) \quad \forall\ \theta \in \Theta$ de donde

$$\lambda_i(R_c(\theta)) = 1-\lambda_i(L_c(\theta)) \quad i=1,\dots,k \quad (2.4.3)$$

Ahora, el teorema 1 nos conduce inmediatamente al resultado.

□

Corolario 6.1

Si $R_c(\theta)$ es una matriz simétrica, entonces $R_c(\theta)$ es semidefinida positiva.

Demostración

Es consecuencia del teorema 3, p. 54 de (Bellman; 1970) y el corolario 6 i).

□

Como caso particular

Corolario 6.2

Si $I_c(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$ entonces

La matriz $R_c(\theta)$ es diagonal, semidefinida positiva con elemento i-ésimo la eficiencia relativa acerca de θ_i (componente i-ésima de θ) para todo $\theta \in \Theta$.

Demostración

Según el corolario 2, la matriz $L_c(\theta)$ es diagonal con elemento i-ésimo

$$L_c^{ii}(\theta) = 1- \frac{I_c(\theta_i)}{I_o(\theta_i)}$$ la pérdida relativa de información acerca de θ_i

Entonces, la matriz $R_c(\theta) = I-L_c(\theta)$ será diagonal con elemento i-ésimo

$$R_c^{ii} = \frac{I_c(\theta_i)}{I_o(\theta_i)}$$ es decir, la eficiencia relativa,

acerca de θ_i, de ε_c comparado con ε_o.

Por último $R_c(\theta) \geq 0$ (Corolario 6.1)

□

La propiedad * 2 tiene su generalización matricial en el siguiente corolario.

Corolario 7

Sean $C_1 \overset{st}{\leq} C_2$ dos variables de censura e $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

i) La matriz $R_{c_2}(\theta)-R_{c_1}(\theta)$ tiene todos sus autovalores no negativos para todo $\theta \in \Theta$

ii) $\lambda_i(R_{c_2}(\theta)) \geq \lambda_i(R_{c_1}(\theta))$ $i=1,\ldots,k$ para todo $\theta \in \Theta$

Demostración

Por definición $R_{C_i}(\theta) = I-L_{C_i}(\theta)$ $i=1,2$, de donde

$$\lambda_j(R_{C_i}(\theta)) = 1-\lambda_j(L_{C_i}(\theta)) \quad j=1,\dots,k \ , \quad i=1,2$$

el teorema 2 completa la demostración.

□

Sea ahora $\phi = f(\theta)$ una transformación biyectiva de θ y $A=(a_{ij})$ con $a_{ij} = \dfrac{\partial\theta_i}{\partial\phi_j}$ $i,j=1,\dots,k$; la propiedad * 3 tiene su equivalente matricial en el siguiente corolario.

Corolario 8

Si $I_C(\theta) > 0$ y A es no singular, para todo $\theta \in \Theta$, entonces

Las matrices $R_C(\theta)$ y $R_C(\phi)$ tienen los mismos autovalores, para toda variable C y $\theta \in \Theta$.

Demostración

Consecuencia inmediata de la expresión (2.4.3) y del teorema 3.

□

Corolario 8.1

Si $I_C(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$, entonces

Los autovalores de la matriz $R_C(\phi)$ son las eficiencias relativas acerca de cada componente θ_i de θ.

Demostración

La matriz $R_C(\theta)$ es diagonal (corolario 6.2) por lo que sus autovalores serán sus elementos diagonales. El corolario 8 completa la demostración.

□

Como consecuencia de este corolario:

Siempre que podamos transformar el parámetro ϕ mediante una transformación biyectiva en otro θ que haga diagonal la matriz de Fisher, la matriz de eficiencias, acerca de θ o de ϕ, del experimento censurado ε_c comparado con el no censurado ε_o, tiene por autovalores las eficiencias relativas acerca de cada componente de θ.

Corolario 8.2

Sean $C_1 \overset{st}{\leqslant} C_2$, $I_o(\theta) > 0$ y A no singular, para todo $\theta \in \Theta$, entonces

Las matrices $R_{C_2}(\theta)-R_{C_1}(\theta)$ y $R_{C_2}(\phi)-R_{C_1}(\phi)$ tienen los mismos autovalores.

Demostración

Por definición $R_{C_2}(\theta)-R_{C_1}(\theta) = L_{C_1}(\theta)-L_{C_2}(\theta)$ y análogamente para ϕ.

El corolario 4 completa la demostración.

□

Para finalizar, estudiamos el comportamiento de la matriz de eficiencias cuando los experimentos observados son $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$. Denotamos por $R_{n,c}(\theta)$ a la matriz de eficiencias acerca de θ del experimento $\varepsilon_c^{(n)}$ comparado con $\varepsilon_o^{(n)}$, es decir

$$R_{n,c}(\theta) = I-L_{n,c}(\theta)$$

La propiedad * 4 tiene su análogo matricial en el siguiente corolario.

Corolario 9

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

$R_{n,c}(\theta) = R_c(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Demostración

Es inmediata del teorema 4.

□

2.4.3. EFICIENCIA RELATIVA. MEDIDAS REALES

Presentamos en este apartado tres medidas reales de la eficiencia relativa que definiremos mediante otras tantas funciones reales de la matriz de eficiencias y más concretamente, de sus autovalores.

Exigiremos a estas medidas que verifiquen, cuando menos, las propiedades * 1 a 4 que exponíamos al comienzo de la sección 2.4 para la medida real de la eficiencia relativa correspondiente a la medida de información de Fisher (unidimensional).

Proponemos las mismas funciones reales de los autovalores que utilizamos para definir las medidas reales de la pérdida relativa de información, allí para la matriz de pérdidas, aquí para la matriz de eficiencias.

Definición 2.4.2

La función real

$$R_c^1(\theta) = \frac{1}{k} \sum_{i=1}^{k} \lambda_i(R_c(\theta))$$

siempre que esté definida,es una medida real de la eficiencia relativa, acerca de θ, del experimento ε_c comparado con ε_o.

La medida $R_c^1(\theta)$ es la media aritmética de los autovalores de la matriz de eficiencias $R_c(\theta)$ y está definida siempre que $I_o(\theta) > 0$ para todo $\theta \in \Theta$.

Una expresión alternativa de $R_c^1(\theta)$ será:

$$R_c^1(\theta) = \frac{1}{k} \text{Traza}[R_c(\theta)]$$

La relación entre las medidas $R_c^1(\theta)$ y $L_c^1(\theta)$ es evidente:

$$\underline{R_c^1(\theta)} = \frac{1}{k} \sum_{i=1}^{k} [1-\lambda_i(L_c(\theta))] = 1- \frac{1}{k} \sum_{i=1}^{k} \lambda_i(L_c(\theta)) =$$

$$= \underline{1-L_c^1(\theta)}$$

Denotaremos por $R_{n,c}^1(\theta)$ a la generalización de la medida real anterior para los experimentos $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$, es decir

$$R_{n,c}^1(\theta) = \frac{1}{k} \text{Traza}[R_{n,c}(\theta)] = 1-L_{n,c}^1(\theta)$$

Debemos probar que la medida real $R_c^1(\theta)$ verifica las propiedades * 1 a 4 ; el próximo corolario recoge este resultado.

Sean $C_1 \overset{st}{\leqslant} C_2$ dos variables de censura y $\phi=f(\theta)$ una transformación biyectiva de θ con $A=(\frac{\partial\theta_i}{\partial\phi_j})$ $i,j=1,\dots,k$

Corolario 10

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

a) $\begin{cases} \text{i)} & 0 \leq R_C^1(\theta) \leq 1 \quad \text{para toda variable } C \text{ y } \theta \in \Theta \\ \text{ii)} & 1 = R_C^1(\theta) \Leftrightarrow I_o(\theta) = I_C(\theta) \\ \text{iii)} & 0 = R_C^1(\theta) \Leftrightarrow I_C(\theta) = \underset{\sim}{0} \end{cases}$

b) $R_{C_2}^1(\theta) \geq R_{C_1}^1(\theta)$ para todo $\theta \in \Theta$

c) $R_{n,C}^1(\theta) = R_C^1(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Si además la matriz A es no singular para todo $\theta \in \Theta$

d) $R_C^1(\theta) = R_C^1(\phi)$ para toda variable C

Demostración

La definición 2.4.2 de $R_C^1(\theta)$ y los corolarios 6, 7 ii), 9 y 8 conducen a los resultados a) b) c) y d) respectivamente.

Una forma alternativa de demostración sería a partir del teorema 5, teniendo en cuenta que $R_C^1(\theta) = 1-L_C^1(\theta)$.

$\square$

Como caso particular, tenemos.

Corolario 10.1

Si $I_C(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$ entonces $R_C^1(\theta)$ es la media aritmética de las eficiencias relativas, acerca de cada θ_i (componente i-ésima de θ) para todo $\theta \in \Theta$

Demostración

Es consecuencia inmediata de la definición 2.4.2 y del corolario 6.2.

$\square$

A continuación, definimos una segunda medida real de la eficiencia relativa.

Definición 2.4.3

La función real

$$R_c^2(\theta) = \left[\prod_{i=1}^{k} \left[1+\lambda_i(R_c(\theta))\right]\right]^{1/k} - 1$$

siempre que esté definida, es una medida real de la eficiencia relativa, acerca de θ, del experimento ε_c comparado con ε_o.

La medida $R_c^2(\theta)$ es la media geométrica de los autovalores de la matriz $I+R_c(\theta)$ menos uno, o también la φ-media de los autovalores de la matriz $R_c(\theta)$, correspondiente a la función $\varphi(x) = \log(1+x)$ (Calot; 1970, p. 73); y está definida, como ya es habitual, siempre que $I_o(\theta) > 0$ para todo $\theta \in \Theta$

Una expresión alternativa de $R_c^2(\theta)$ será:

$$R_c^2(\theta) = \left|I+R_c(\theta)\right|^{1/k} - 1$$

Como es de esperar, $R_{n,c}^2(\theta)$ denota la medida real anterior para los experimentos $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$, es decir:

$$R_{n,c}^2(\theta) = \left|I+R_{n,c}(\theta)\right|^{1/k} - 1$$

Veamos que la medida $R_c^2(\theta)$ verifica las propiedades * 1 a 4.

Sea $C_1 \overset{st}{\leq} C_2$ y $\phi=f(\theta)$ una transformación biyectiva de θ con $A = (\frac{\partial\theta_i}{\partial\varphi_j})$ $i,j=1,\ldots,k$.

Corolario 11

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

a) $\begin{cases} \text{i)} & 0 \leqslant R_c^2(\theta) \leqslant 1 \quad \text{para toda variable } C \text{ y } \theta \in \Theta \\ \text{ii)} & 1 = R_c^2(\theta) \Leftrightarrow I_c(\theta) = I_o(\theta) \\ \text{iii)} & 0 = R_c^2(\theta) \Leftrightarrow I_c(\theta) = \underset{\sim}{0} \end{cases}$

b) $R_{c_2}^2(\theta) \geqslant R_{c_1}^2(\theta)$ para todo $\theta \in \Theta$

c) $R_{n,c}^2(\theta) = R_c^2(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Si además la matriz A es no singular para todo $\theta \in \Theta$

d) $R_c^2(\theta) = R_c^2(\phi)$ para toda variable C.

Demostración

La definición 2.4.3 de $R_c^2(\theta)$ y los corolarios 6, 7 ii), 9 y 8 conducen a los resultados a) b) c) y d) respectivamente.

$\square$

Definimos, por último, una tercera medida real de la eficiencia relativa.

Definición 2.4.4

La función real

$$R_c^3(\theta) = \left[\frac{\sum_{i=1}^{k} \left[\lambda_i(R_c(\theta))\right]^2}{k} \right]^{1/2}$$

seimpre que esté definida, es una medida real de la eficiencia relativa, acerca de θ, del experimento ε_c comparado con ε_o.

La medida $R_c^3(\theta)$ es la media cuadrática de los autovalores de la matriz de eficiencias $R_c(\theta)$ y como siempre, su definición requiere que $I_o(\theta) > 0$ para todo $\theta \in \Theta$

En ocasiones, la matriz de eficiencias es diagonal (como ocurre en la aplicación práctica del capítulo III), en este caso y en general si la matriz de eficiencias es simétrica, la medida real $R_c^3(\theta)$ es la norma euclídea de la matriz $\frac{1}{\sqrt{k}} R_c(\theta)$.

En efecto:

$$R_c^3(\theta) = \frac{1}{\sqrt{k}} \left[\sum_{i=1}^{k} \left[\lambda_i(R_c(\theta))\right]^2 \right]^{1/2} = \frac{1}{\sqrt{k}} \| R_c(\theta) \| = \| \frac{1}{\sqrt{k}} R_c(\theta) \|$$

Es sencillo comprobar que esta medida verifica las propiedades * 1 a 4.

A partir de la expresión (2.4.3) podemos escribir:

$$R_c^3(\theta) = \left[1+(L_c^3(\theta))^2-2L_c^1(\theta)\right]^{1/2}$$

Terminamos esta sección 2.4. con algunas observaciones relativas a la matriz de eficiencias R_c y a sus medidas reales - asociadas R_c^i i=1,2,3.

<u>Observaciones</u>

1.- La matriz

$R_c' = I-L_c'$ sería otra medida matricial de la eficiencia relativa, con los mismos autovalres que R_c, de forma que todas las definiciones y propiedades que hemos vis to hacen indistinta la elección de R_c o R'_c como medida matricial de la eficiencia relativa.

2.- Por definición, las medidas reales R_c^i $i=1,2,3$ son tres promedios de los autovalores de la matriz de eficiencias R_c, por lo que podemos escribir que

$$\lambda_k(R_c) \leq R_c^i \leq \lambda_1(R_c) \qquad \text{para toda } C, \quad i=1,2,3$$

Además la relación entre las medias geométrica, aritmética y cuadrática (Calot; 1970, pp. 73-77) hace que

$$R_c^2(\theta) \leq R_c^1(\theta) \leq R_c^3(\theta) \qquad \text{para toda } C \text{ y } \theta \in \Theta$$

dándose las dos igualdades a la vez, si y sólo si, para θ y C fijos, todos los autovalores de la matriz de eficiencias coinciden.

La discusión, en cuanto a la elección de una de estas tres medidas, se reducirá pues, en cada caso, al análisis de la representatividad, ventajas e inconvenientes de estos promedios.

3.- Si la matriz de Fisher $I_c(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$ entonces la matriz de eficiencias R_c también es diagonal y las medidas reales R_c^i $i=1,2,3$ son los promedios correspondientes de las eficiencias relativas, acerca de cada θ_i (componente i-ésima de θ) para todo $\theta \in \Theta$; R_c^1 la media aritmética, R_c^2 la φ-media, con $\varphi(x)=\log(1+x)$ y R_c^3 la media cuadrática.

4.- Brooks (1982) definía la eficiencia relativa límite de un experimento censurado respecto al no censurado, como el límite, para $n\to\infty$, de la eficiencia relativa de $\varepsilon_C^{(n)}$ respecto de $\varepsilon_O^{(n)}$.

Si aplicamos esta definición a las medidas reales R_C^i i=1,2,3, por la propiedad * 3 de dichas medidas tenemos que

$$\lim_{n\to\infty} R^i_{n,C}(\theta) = R^i_C(\theta) \qquad \text{para toda variable C y } \theta \in \Theta$$
$$i=1,2,3$$

es decir:

La eficiencia relativa límite es la eficiencia relativa de una observación del experimento censurado respecto del no censurado.

2.5. MEDIDAS REALES

Al comienzo de este capítulo, anunciábamos dos maneras de evaluar la pérdida relativa de información y la eficiencia relativa en un experimento censurado, cuando la medida de información adecuada es la matriz de Fisher.

La primera ha sido descrita a lo largo de las secciones 2.3 y 2.4 y consiste, en síntesis, en definir medidas matriciales de la pérdida relativa de información y de la eficiencia relativa, para posteriormente definir medidas reales de las mismas a través de las medidas matriciales anteriores. También veíamos que todo este proceso podía resultar estéril, si la matriz

I_o fuese singular.

En este caso podríamos abordar nuestro objetivo de una segunda forma que exponemos en esta sección y que consiste, en síntesis, en definir medidas de información reales a partir de la matriz de Fisher y calcular posteriormente, la pérdida relativa de información y la eficiencia relativa reales de igual forma que lo hace Brooks (1982). Utilizando, como medidas reales de información las funciones reales de la matriz de Fisher, S_X, D_X y M_X, propuestas en el capítulo I, dedicaremos los próximos apartados al análisis de las propiedades de las medidas reales de la pérdida relativa de información y de la eficiencia relativa, inducidas por aquéllas.

2.5.1. MEDIDA DE INFORMACION S_c

Denotamos por $S_c(\theta)$ a la medida de información parámetrica, definida en el epígrafe 1.3.3.1, relativa al experimento ε_c, es decir

$$S_c(\theta) = \alpha \sum_{i=1}^{k} \lambda_i(I_c(\theta)) \qquad \text{para todo } \alpha>0 \text{ fijo}$$

De forma análoga denotaremos por $S_o(\theta)$, $S_{n,c}(\theta)$ y $S_{n,o}(\theta)$ a la medida de información anterior relativa a los experimentos $\varepsilon_o, \varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$ respectivamente.

2.5.1.1. PERDIDA RELATIVA DE INFORMACION

Si elegimos como medida de información, la medida paramétrica S_c, la pérdida relativa de información, acerca de θ, al observar ε_c en lugar de ε_o, que denotamos por $L_c^S(\theta)$, será el número real

$$L_c^S(\theta) = \frac{S_o(\theta)-S_c(\theta)}{S_o(\theta)} = 1- \frac{S_c(\theta)}{S_o(\theta)} \qquad (2.5.1)$$

Sabemos que $S_o(\theta) > 0$ salvo que $f(x,\theta)$ no dependa de θ (propiedad (1) de S_c). En lo que sigue, supondremos pues que $S_o(\theta) > 0$ para todo $\theta \in \Theta$

Si tenemos en cuenta que $S_c(\theta) = \alpha \sum_{i=1}^{k} I_c(\theta_i)$ para toda C entonces

$$L_c^S(\theta) = 1- \frac{\sum_{i=1}^{k} I_c(\theta_i)}{\sum_{i=1}^{k} I_o(\theta_i)} = \sum_{i=1}^{k}\left[1- \frac{I_c(\theta_i)}{I_o(\theta_i)}\right] \frac{I_o(\theta_i)}{\sum_{i=1}^{k} I_o(\theta_i)} \qquad (2.5.2)$$

es decir:

$L_c^S(\theta)$ es la media aritmética ponderada de las pérdidas relativas de información acerca de cada θ_i, $i=1,\dots,k$, siendo las ponderaciones respectivas las proporciones de información de Fisher acerca de cada θ_i según el experimento ε_o.

Como ya es habitual, $L_{C_i}^S$ denotará la medida anterior para el experimento censurado ε_{C_i} según la variable de censura C_i, y $L_{n,c}^S$ expresará la generalización de dicha medida para los experimentos $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$.

El próximo teorema recoge las principales propiedades de la medida $L_c^S(\theta)$. Sean $C_1 \overset{st}{\leq} C_2$ dos variables de censura.

<u>Teorema 7</u>

a) $\begin{cases} \text{i)} & 0 \leq L_c^S(\theta) \leq 1 \quad \text{para toda variable C y } \theta \in \Theta \\ \text{ii)} & 0 = L_c^S(\theta) \Leftrightarrow S_c(\theta) = S_o(\theta) \Leftrightarrow I_c(\theta) = I_o(\theta) \end{cases}$

iii) $1 = L_c^S(\theta) \Leftrightarrow S_c(\theta) = 0 \Leftrightarrow I_c(\theta) = \underset{\sim}{0}$

b) $L_{c_1}^S(\theta) \geq L_{c_2}^S(\theta)$ para todo $\theta \in \Theta$

c) $L_{n,c}^S(\theta) = L_c^S(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Demostración

a) $\varepsilon_o \gtrsim \varepsilon_c$ para toda variable de censura C (Goel;1987,cor.3.2)

entonces

$$S_c(\theta) \leq S_o(\theta) \quad \text{(propiedad (5) de la medida } S_c)$$

además sabemos que

$S_c(\theta) = S_o(\theta) \Leftrightarrow \text{Traza}\left[I_o(\theta)-I_c(\theta)\right] = 0 \Leftrightarrow \lambda_i(I_o(\theta)-I_c(\theta)) = 0 \Leftrightarrow I_c(\theta) = I_o(\theta)$ $\quad i=1,\ldots,k$

De otra parte

$0 \leq S_c(\theta)$ y $S_c(\theta) = 0 \Leftrightarrow I_c(\theta) = \underset{\sim}{0}$ (propiedad (1) de S_c)

la expresión (2.5.1) y el hecho de que $S_o(\theta) > 0$ completan la demostración.

□

b) $\varepsilon_{c_1} \lesssim \varepsilon_{c_2}$ por ser $C_1 \overset{st}{\leq} C_2$ (Goel; 1987, th. 3.1)

entonces

$S_{c_1}(\theta) \leq S_{c_2}(\theta)$ (propiedad (5) de la medida S_c)

que conduce inmediatamente al resultado.

c) Es consecuencia inmediata de la propiedad (2) de aditividad para observaciones independientes de la medida S_c.

□

La medida L_c^S no verifica la propiedad de invariancia para cualquier transformación biyectiva del espacio paramétrico. Si $\phi=f(\theta)$ es una transformación biyectiva de θ, la expresión (2.5.1) exige la siguiente equivalencia

$$L_c^S(\theta) = L_c^S(\phi) \Leftrightarrow \frac{S_c(\theta)}{S_o(\theta)} = \frac{S_c(\phi)}{S_o(\phi)}$$

La definición de la medida S_c hace que esta última igualdad sea equivalente a la siguiente

$$\frac{\text{Traza}(I_c(\theta)AA')}{\text{Traza}(I_o(\theta)AA')} = \frac{\text{Traza}(I_c(\theta))}{\text{Traza}(I_o(\theta))} \quad \text{con } A=(\frac{\partial\theta_i}{\partial\phi_j}) \qquad i,j=1,\ldots,k$$

igualdad que sólo podemos asegurar en el caso de que ϕ sea una transformación lineal ortogonal (ver propiedades 1 y 3 epígrafe 1.3.4.1)

2.5.1.2. EFICIENCIA RELATIVA

Denotamos por $R_{n,c}^S(\theta)$ al número real que mide la eficiencia relativa acerca de θ, del experimento $\varepsilon_c^{(n)}$ comparado con $\varepsilon_o^{(n)}$, cuando la medida de información utilizada es la medida paramétrica S_c. Para los experimentos ε_c y ε_o usaremos la notación $R_c^S(\theta)$. Siguiendo a Brooks (1982):

$$R_{n,c}^S(\theta) = \frac{n^*(\theta)}{n} \qquad \text{siendo } n^* \in R \;,, \; S_{n^*,o}(\theta) = S_{n,c}(\theta)$$

Por ser la medida S_c aditiva para observaciones independientes, tenemos que

$$S_{n^*,o}(\theta) = S_{n,c}(\theta) \Leftrightarrow n^*S_o(\theta) = nS_c(\theta) \qquad \text{entonces}$$

$$R^S_{n,c}(\theta) = \frac{S_c(\theta)}{S_o(\theta)} = R^S_c(\theta) \qquad \text{para todo } n \in \mathbb{N} \qquad (2.5.3)$$

resultado que ya fue puesto de manifiesto para cualesquiera medidas paramétricas de información con la propiedad de aditividad (ver (2.4.1)).

Teniendo en cuenta las expresiones (2.5.1) y (2.5.2)

$$R^S_c(\theta) = 1-L^S_c(\theta) \qquad \text{para toda variable C y } \theta \in \Theta \qquad (2.5.4)$$

y

$$R^S_c = \sum_{i=1}^{k} \left[\frac{I_c(\theta_i)}{I_o(\theta_i)}\right] \frac{I_o(\theta_i)}{\sum\limits_{i=1}^{k} I_o(\theta_i)} \qquad (2.5.5)$$

es decir

R^S_c es la media aritmética ponderada de las eficiencias relativas acerca de cada θ_i, $i=1,\ldots,k$, siendo las ponderaciones respectivas las proporciones de información de Fisher acerca de cada θ_i según el experimento ε_o.

El siguiente corolario recoge las propiedades de la medida $R^S_c(\theta)$. Suponemos, como siempre, $C_1 \overset{st}{\leqslant} C_2$ dos variables de censura.

Corolario 12

a)
- i) $0 \leqslant R^S_c(\theta) \leqslant 1$ para toda variable C y $\theta \in \Theta$
- ii) $1 = R^S_c(\theta) \Leftrightarrow S_c(\theta) = S_o(\theta) \Leftrightarrow I_c(\theta) = I_o(\theta)$
- iii) $0 = R^S_c(\theta) \Leftrightarrow S_c(\theta) = 0 \Leftrightarrow I_c(\theta) = \underset{\sim}{0}$

b) $R^S_{c_2}(\theta) \geqslant R^S_{c_1}(\theta)$ para todo $\theta \in \Theta$

c) $R^S_{n,c}(\theta) = R^S_c$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Demostración

a) y b) son inmediatas, a partir de la expresión (2.5.4) y del teorema 7 a) y b). La expresión (2.5.3) recoge la parte c) del corolario.

□

La estrecha relación existente entre las medidas L_c^S y R_c^S hace que R_c^S no sea invariante (para cualesquiera transformaciones biyectivas del parámetro) al no serlo L_c^S.

Observación

Llamamos la atención sobre el parecido, que no igualdad, de las medidas R_c^1 y R_c^S y como consecuencia, el de L_c^1 y L_c^S

$$R_c^1 = \frac{\text{Traza}(I_c I_o^{-1})}{k} \qquad R_c^S = \frac{\text{Traza}(I_c)}{\text{Traza}(I_o)}$$

Si la matriz de Fisher $I_c(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$, una condición suficiente para que ambas medidas R_c^1 y R_c^S (L_c^1 y L_c^S) coincidan en $\theta \in \Theta$ es que

$$\frac{I_o(\theta_i)}{\sum_{i=1}^{k} I_o(\theta_i)} = \frac{1}{k} \qquad i=1,\ldots,k$$

2.5.2. MEDIDA DE INFORMACION D_c

Denotamos por $D_c(\theta)$ a la medida de información paramétrica, definida en el epígrafe 1.3.3.2, relativa al experimento ε_c, es decir

$$D_c(\theta) = \left[\prod_{i=1}^{k} \lambda_i(I_c(\theta))\right]^{\alpha} \qquad \text{para } \alpha > 0 \text{ fijo}$$

De forma análoga, denotaremos por $D_o(\theta)$, $D_{n,c}(\theta)$ y $D_{n,o}(\theta)$ a la medida de información anterior relativa a los experimentos ε_o, $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$ respectivamente.

2.5.2.1. PERDIDA RELATIVA DE INFORMACION

Si elegimos como medida de información la medida paramétrica D_c, la pérdida relativa de información, acerca de θ, al observar ε_c en lugar de ε_o, que denotamos por $L_c^D(\theta)$, será el número real

$$L_c^D(\theta) = \frac{D_o(\theta)-D_c(\theta)}{D_o(\theta)} = 1- \frac{D_c(\theta)}{D_o(\theta)} \qquad (2.5.6)$$

La definición de esta medida requiere que $D_o(\theta) > 0$ para todo $\theta \in \Theta$ lo que exige que la matriz $I_o(\theta)$ sea no singular para todo $\theta \in \Theta$

Denotaremos, como es habitual, por $L_{c_i}^D$ a la medida anterior para el experimento censurado ε_{c_i} (según la variable de censura C_i) y por $L_{n,c}^D$ a la generalización de esta medida para los experimentos $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$.

El siguiente teorema recoge las principales propiedades de la medida $L_c^D(\theta)$. Sean $C_1 \overset{st}{\leq} C_2$ dos variables de censura y $\phi=f(\theta)$ una transformación biyectiva de θ con $A=(\frac{\partial\theta_i}{\partial\phi_j})$ i,j=1,...,k

Teorema 8

Si $I_o(\theta) > 0$ para todo $\theta \in \Theta$ entonces

a)
- i) $0 \leq L_c^D(\theta) \leq 1$ para toda variable C y $\theta \in \Theta$
- ii) $0 = L_c^D(\theta) \Leftrightarrow D_c(\theta) = D_o(\theta) \Leftarrow I_c(\theta) = I_o(\theta)$
- iii) $1 = L_c^D(\theta) \Leftrightarrow D_c(\theta) = 0 \Leftarrow I_c(\theta) = \underset{\sim}{0}$

b) $L_{c_1}^D(\theta) \geq L_{c_2}^D(\theta)$ para todo $\theta \in \Theta$

c) $L^D_{n,c}(\theta) = L^D_c(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Si además la matriz A es no singular para todo $\theta \in \Theta$

d) $L^D_c(\theta) = L^D_c(\phi)$ para toda variable C.

Demostración

a) El corolario 3.2 de Goel (1987) y las propiedades (1) y (5) de la medida D_c, junto con la expresión (2.5.6) y el hecho de que $D_o(\theta) > 0$ completan la demostración.

b) Es consecuencia del teorema 3.1 de Goel (1987) y de la propiedad (5) de la medida D_c.

c) La propiedad (3) de multiplicidad de la medida D_c y la expresión (2.5.6) nos conducen al resultado.

d) La propiedad 4 del capítulo I (epígrafe 1.3.4.2) y (2.5.6) nos llevan al resultado.

$\square$

2.5.2.2. EFICIENCIA RELATIVA

Demotamos por $R^D_{n,c}(\theta)$ al número real que mide la eficiencia relativa, acerca de θ, del experimento $\varepsilon^{(n)}_c$ comparado con $\varepsilon^{(n)}_o$, cuando utilizamos como medida de información la medida paramétrica D_c. Denotaremos $R^D_c(\theta)$ a dicha eficiencia relativa, para los experimentos ε_c y ε_o.

La medida D_c no es aditiva para observaciones independien tes. Según la definición de Brooks (1982):

$R^D_{n,c}(\theta) = \dfrac{n^*(\theta)}{n}$ siendo, para cada n, $n^* \in \mathbb{R}_+$,

$D_{n^*,o}(\theta) = D_{n,c}(\theta)$

Pero

$D_{n^*,o}(\theta) = (n^*)^{k\alpha} D_o(\theta)$ y análogamente

$D_{n,c}(\theta) = n^{k\alpha} D_c(\theta)$ por la propiedad (3) de multiplicidad de la medida D_c.

Por lo tanto

$$D_{n^*,o}(\theta) = D_{n,c}(\theta) \Leftrightarrow \frac{n^*}{n} = \left[\frac{D_c(\theta)}{D_o(\theta)}\right]^{1/k\alpha}$$

y

$$R^D_{n,c}(\theta) = \left[\frac{D_c(\theta)}{D_o(\theta)}\right]^{1/k\alpha} = R^D_c(\theta) \quad \text{para todo } n \in \mathbb{N} \quad (2.5.7)$$

Teniendo en cuenta la expresión (2.5.6) podemos escribir

$$R^D_c(\theta) = \left[1-L^D_c(\theta)\right]^{1/k\alpha} \qquad (2.5.8)$$

El siguiente corolario recoge las propiedades más importantes de la medida $R^D_c(\theta)$. En la notación habitual, sean $C_1 \overset{st}{\leqslant} C_2$ dos variables de censura y $A = \left(\dfrac{\partial\theta_i}{\partial\varphi_j}\right)$ i,j=1,...,k, siendo $\varphi=f(\theta)$ una transformación biyectiva de θ.

Corolario 13

Si $I_o(\theta)>0$ para todo $\theta \in \Theta$ entonces

$$a)\begin{cases} \text{i)} & 0 \leqslant R^D_c(\theta) \leqslant 1 \quad \text{para toda variable C y } \theta \in \Theta \\ \text{ii)} & 1 = R^D_c(\theta) \Leftrightarrow D_c(\theta) = D_o(\theta) \Leftarrow I_c(\theta) = I_o(\theta) \\ \text{iii)} & 0 = R^D_c(\theta) \Leftrightarrow D_c(\theta) = 0 \Leftarrow I_c(\theta) = \underset{\sim}{0} \end{cases}$$

b) $R^D_{C_2}(\theta) \geq R^D_{C_1}(\theta)$ para todo $\theta \in \Theta$

c) $R^D_{n,C}(\theta) = R^D_C(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Si además la matriz A es no singular para todo $\theta \in \Theta$

d) $R^D_C(\theta) = R^D_C(\phi)$ para toda variable C.

<u>Demostración</u>

a) b) y d) son inmediatas, a partir de la expresión (2.5.8) y del teorema 8 a) b) y d) respectivamente.

La expresión (2.5.7) recoge la parte c) del corolario.

$\square$

Terminamos el apartado 2.5.2 con algunas observaciones interesantes sobre las medidas L^D_C y R^D_C.

<u>Observaciones</u>

1.- Teniendo en cuenta que $D_C(\theta) = \left|I_C(\theta)\right|^{\alpha}$ para toda C, si la matriz de Fisher $I_C(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$ tenemos que

$$R^D_C(\theta) = \left[\prod_{i=1}^{k} \frac{I_C(\theta_i)}{I_o(\theta_i)}\right]^{1/k}$$

es decir

$R^D_C(\theta)$ es la media geométrica de las eficiencias relativas acerca de cada θ_i, $i=1,\ldots,k$ y, eligiendo $\alpha = \frac{1}{k}$,

$$L^D_C(\theta) = 1- \left[\prod_{i=1}^{k} \frac{I_C(\theta_i)}{I_o(\theta_i)}\right]^{1/k}$$

es decir

$L_C^D(\theta)$ es la φ-media de las pérdidas relativas de información acerca de cada θ_i, $i=1,\ldots,k$, siendo $\varphi(x)=\log(1-x)$

2.- Supongamos, de nuevo, que la matriz de Fisher $I_C(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$ y elijamos $\alpha = \frac{1}{k}$, entonces

$$L_C^D(\theta) \geqslant L_C^2(\theta) \qquad \text{para toda C y } \theta \in \Theta$$

con la igualdad si y sólo si para C y θ fijos, coinciden todas las pérdidas relativas de información acerca de θ_i, $i=1,\ldots,k$.

Este resultado se basa en la observación 2 de la sección 2.3 y en la desigualdad

$$L_C^D(\theta) \geqslant L_C^1(\theta) \qquad \text{para toda C y } \theta \in \Theta$$

que se obtiene fácilmente a partir de la relación entre las medias aritmética y geométrica.

3.- La definición de las medidas L_C^D y R_C^D requiere que la matriz $I_O(\theta)$ sea no singular para todo $\theta \in \Theta$. Sin embargo esta hipótesis no garantiza el buen comportamiento de las medidas L_C^D y R_C^D ya que puede ocurrir que

$$L_C^D(\theta) = 0 \quad \text{y ser} \quad I_C(\theta) \neq I_O(\theta)$$

o (teorema 8 a)

$$L_C^D(\theta) = 1 \quad \text{y la matriz } I_C(\theta) \neq \underset{\sim}{0}$$

y análogamente para R_C^D (corolario 13 a)).

Estas lagunas son el reflejo de las correspondientes de la medida D_c y se solventan cuando la matriz $I_c(\theta)$ es no singular para todo $\theta \in \Theta$. Sólo en este caso debemos utilizar las medidas L_c^D y R_c^D

2.5.3. MEDIDA DE INFORMACION M_c

Denotamos por $M_c(\theta)$ a la medida de información paramétrica, definida en el epígrafe 1.3.3.3, relativa al experimento ε_c, es decir:

$$M_c(\theta) = \left[\sum_{i=1}^{k} \left[\lambda_i(I_c(\theta))\right]^2\right]^{1/2}$$

Análogamente $M_o(\theta)$ expresará la medida de información anterior, relativa al experimento ε_o.

2.5.3.1. PERDIDA RELATIVA DE INFORMACION

Si elegimos como medida de información la medida paramétrica M_c, la pérdida relativa de información, acerca de θ, al observar ε_c, en lugar de ε_o, que denotamos por $M_c(\theta)$, será el número real

$$L_c^M(\theta) = \frac{M_o(\theta)-M_c(\theta)}{M_o(\theta)} = 1- \frac{M_c(\theta)}{M_o(\theta)} \qquad (2.5.9)$$

Por la propiedad (1) de la medida M_c sabemos que $M_o(\theta) > 0$, salvo que $f(x,\theta)$ no dependa de θ. Supondremos pues, en lo que sigue que $M_o(\theta) > 0$ para todo $\theta \in \Theta$

Como siempre $L_{c_i}^M$ denotará la medida anterior para ε_{c_i} (experimento censurado según la variable de censura C_i) y $L_{n,c}^M$ expresará la generalización de esta medida para los experimentos $\varepsilon_c^{(n)}$ y $\varepsilon_o^{(n)}$.

El próximo teorema recoge las principales propiedades de la medida $L_c^M(\theta)$. Sean $C_1 \overset{st}{\leq} C_2$ dos variables de censura.

<u>Teorema 9</u>

a)
- i) $0 \leq L_c^M(\theta) \leq 1$ para toda variable C y $\theta \in \Theta$
- ii) $0 = L_c^M(\theta) \Leftrightarrow M_c(\theta) = M_o(\theta) \Leftrightarrow I_c(\theta) = I_o(\theta)$
- iii) $1 = L_c^M(\theta) \Leftrightarrow M_c(\theta) = 0 \Leftrightarrow I_c(\theta) = \underset{\sim}{0}$

b) $L_{c_1}^M(\theta) \geq L_{c_2}^M(\theta)$ para todo $\theta \in \Theta$

c) $L_{n,c}^M(\theta) = L_c^M(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

<u>Demostración</u>

a) El corolario 3.2 de Goel (1987) y las propiedades (1) y (6) de la medida M_c, junto con la expresión (2.5.9) y el hecho de que $M_o(\theta) > 0$ completan la demostración.

b) Es consecuencia del teorema 3.1 de Goel (1987) y de la propiedad (6) de la medida M_c.

c) La propiedad (3) de aditividad para observaciones independientes de la medida M_c y la expresión (2.5.9) nos conducen al resultado.

□

Al igual que ocurría con L_c^S, la propiedad de invariancia para cualquier transformación biyectiva del parámetro no se cumple para la medida L_c^M. Si $\phi=f(\theta)$ es una transformación biyectiva del parámetro θ, la expresión (2.5.9) exige la siguiente equivalencia

$$L_c^M(\theta) = L_c^M(\phi) \Leftrightarrow \frac{M_c(\theta)}{M_o(\theta)} = \frac{M_c(\phi)}{M_o(\phi)}$$

La definición de la medida M_c hace que podamos expresar esta última igualdad como

$$\frac{||I_c(\theta)||}{||I_o(\theta)||} = \frac{||I_c(\phi)||}{||I_o(\phi)||}$$

igualdad, que sólo podemos asegurar en el caso de que ϕ sea una transformación lineal ortogonal (Ver propiedades 5 y 6 epígrafe 1.3.4.3).

2.5.3.2. EFICIENCIA RELATIVA

Denotamos, como es habitual, por $R_{n,c}^M(\theta)$ al número real que mide la eficiencia relativa, acerca de θ, del experimento $\varepsilon_c^{(n)}$ comparado con $\varepsilon_o^{(n)}$, cuando utilizamos como medida de información la medida paramétrica M_c. Asimismo $R_c^M(\theta)$ expresará dicha eficiencia relativa, para los experimentos ε_c y ε_o.

Por ser la medida M_c aditiva para observaciones indepen--

tes, podemos escribir

$$R^M_{n,c}(\theta) = \frac{M_c(\theta)}{M_o(\theta)} = R^M_c(\theta) \quad \text{para todo } n \in \mathbb{N} \qquad (2.5.10)$$

y (teniendo en cuenta 2.5.9):

$$R^M_c(\theta) = 1-L^M_c(\theta) \quad \text{para toda variable C y } \theta \in \Theta \qquad (2.5.11)$$

Las propiedades de la medida R^M_c quedan reflejadas en el siguiente corolario. Sean $C_1 \overset{st}{\leq} C_2$ dos variables de censura

Corolario 14

a) $\begin{cases} \text{i)} & 0 \leq R^M_c(\theta) \leq 1 \quad \text{para toda variable C y } \theta \in \Theta \\ \text{ii)} & 1 = R^M_c(\theta) \Leftrightarrow M_c(\theta) = M_o(\theta) \Leftrightarrow I_c(\theta) = I_o(\theta) \\ \text{iii)} & 0 = R^M_c(\theta) \Leftrightarrow M_c(\theta) = 0 \Leftrightarrow I_c(\theta) = \underset{\sim}{0} \end{cases}$

b) $R^M_{c_2}(\theta) \geq R^M_{c_1}(\theta)$ para todo $\theta \in \Theta$

c) $R^M_{n,c}(\theta) = R^M_c(\theta)$ para todo $n \in \mathbb{N}$, toda variable C y $\theta \in \Theta$

Demostración

a) y b) son inmediatas de la expresión (2.5.11)y del teorema 9 a) y b). La expresión (2.5.10)recoge la parte c) del corolario.

□

La relación (2.5.11)hace que la medida R^M_c no sea invariante (para cualesquiera transformaciones biyectivas del parámetro) al no serlo L^M_c. Las transformaciones lineales ortogonales del parámetro dejan invariante la medida R^M_c.

Terminamos el apartado 2.5.3 con algunas observaciones acerca de las medidas L^M_c y R^M_c .

Observaciones

1.- Si la matriz de Fisher $I_c(\theta)$ es diagonal para toda variable C y $\theta \in \Theta$, por la definición de la medida real M_c tenemos que

$$M_c(\theta) = \left[\sum_{i=1}^{k} \left[I_c(\theta_i)\right]^2\right]^{1/2} \quad \text{para toda C}$$

por lo que la expresión (2.5.10) implica que

$$R_c^M(\theta) = \left[\frac{\sum_{i=1}^{k} \left[I_c(\theta_i)\right]^2}{\sum_{i=1}^{k} \left[I_o(\theta_i)\right]^2}\right]^{1/2}$$

$$= \left[\sum_{i=1}^{k} \left[\frac{I_c(\theta_i)}{I_o(\theta_i)}\right]^2 \frac{\left[I_o(\theta_i)\right]^2}{\sum_{i=1}^{k} \left[I_o(\theta_i)\right]^2}\right]^{1/2}$$

es decir

$R_c^M(\theta)$ es la media cuadrática ponderada de las eficiencias relativas acerca de cada θ_i, $i=1,\ldots,k$, siendo las ponderaciones respectivas las proporciones de las informaciones de Fisher al cuadrado, acerca de cada θ_i, según el experimento ε_o.

La expresión (2.5.11) hace que

$$L_c^M(\theta) = 1- \left[\sum_{i=1}^{k} (1-x_i)^2 f_i\right]^{1/2} .$$

con

$$\begin{cases} x_i = 1 - \dfrac{I_c(\theta_i)}{I_o(\theta_i)} \\[2ex] f_i = \dfrac{\left[I_o(\theta_i)\right]^2}{\sum\limits_{i=1}^{k} \left[I_o(\theta_i)\right]^2} \end{cases} \qquad i=1,\ldots,k$$

es decir

$L_c^M(\theta)$ es la φ-media ponderada de las pérdidas relativas de información acerca de cada θ_i, $i=1,\ldots,k$, con las mismas ponderaciones que antes, y $\varphi(x) = (1-x)^2$.

2.- Si la matriz de Fisher $I_c(\theta)$ es diagonal, para toda C y $\theta \in \Theta$, una condición suficiente para que las medidas R_c^M y R_c^3 coincidan en $\theta \in \Theta$ es que

$$\frac{\left[I_o(\theta_i)\right]^2}{\sum\limits_{i=1}^{k} \left[I_o(\theta_i)\right]^2} = \frac{1}{k} \qquad i=1,\ldots,k$$

Para completar la sección 2.5, añadimos, por último, un resultado ya esperado para las tres medidas de la eficiencia relativa

$$\lim_{n\to\infty} R_{n,c}^H(\theta) = R_c^H(\theta) \qquad \text{para toda variable C y } \theta \in \Theta$$

$$H=S,D,M$$

es decir:

la eficiencia relativa límite es la eficiencia relativa de una observación del experimento censurado respecto del no censurado.

CAPITULO III

APLICACION A UN MODELO MULTINOMIAL

CAPITULO III

APLICACION A UN MODELO MULTINOMIAL

3.0. SUMARIO

El objetivo de este capítulo es la aplicación a un caso práctico de los conceptos teóricos propuestos en los dos capítulos precedentes. Para ello introducimos de una manera intuitiva un modelo de supervivencia censurado por la derecha (Sección 3.1) cuya formalización induce una familia de experimentos ε_c en los que observamos una variable aleatoria de distribución multinomial, cuya dimensión depende de la distribución prefijada c de la variable de censura y cuya ley de probabilidad depende de un parámetro k-variante p, para toda c. El experimento no censurado correspondiente ε_o se convierte en un caso particular de ε_c, cuando la variable de censura es degenerada en $+\infty$ (Sección 3.2).

Como resultado previo probamos que el experimento ε_o es suficiente para el experimento ε_c, para toda c, según el concepto de suficiencia de Blackwell (1951, 1953) (Sección 3.3).

Calculamos las medidas paramétricas de información, acerca de p, adecuadas: la matriz de Fisher $I_c(p)$ (Sección 3.4) y las medidas reales $S_c(p)$, $D_c(p)$ y $M_c(p)$, basadas en dicha matriz (Sección 3.6), que proponíamos en el capítulo I. Con objeto de facilitar el cálculo de dichas medidas, definimos previamente una reparametrización q, que tiene la propiedad de hacer diagonal la matriz de Fisher $I_c(q)$. Debido a la relación entre ambas matrices $I_c(p)$ e $I_c(q)$ (Sección 3.5), todas las medidas de información acerca de p pueden obtenerse a partir de esta última matriz $I_c(q)$ (mucho más simple que $I_c(p)$).

Todas estas medidas de información verifican la propiedad de suficiencia de experimentos, lo que se traduce en una pérdida de información, acerca del parámetro, al observar ε_c en vez de ε_o. Evaluaremos esta pérdida de información siguiendo los dos métodos propuestos en el capítulo II: el método matricial (Sección 3.7) y el método real (Sección 3.8). El primero incluye el cálculo de la matriz de Pérdidas (Apartado 3.7.1) y de la de Eficiencias (Apartado 3.7.2) respecto de q, que resultarán ser diagonales. A partir de estas matrices calculamos las medidas reales de la pérdida relativa de información y de la eficiencia relativa, respectivamente. Como resultado más llamativo, observamos que ninguna de dichas medidas depende de q.

La propiedad de invariancia de estas medidas, para transformaciones biyectivas del parámetro, hace innecesario el cálculo de las mismas respecto de p.

El método real incluye, igualmente, el cálculo de las medidas de la pérdida relativa de información y de la eficiencia relativa respecto de p y q para cada una de las medidas de información S_c, D_c y M_c (Apartados 3.8.1, 3.8.2 y 3.8.3 respectivamente). Ahora, todas las medidas, excepto las derivadas de la medida de información D_c, son funciones del parámetro considerado.

3.1. INTRODUCCION

Sea T la variable aleatoria no negativa que representa el tiempo de vida en un estudio de supervivencia, siendo F su función de distribución que supondremos absolutamente continua.

Consideramos una partición del tiempo $(0,\infty) = \bigcup_{i=1}^{k+1} (t_{i-1},t_i]$ con $t_o=0$ y $t_{k+1}=\infty$. Por restricciones de observación, el investigador sólo puede observar la variable en los instantes $t_1 \leq t_2 \leq \ldots \leq t_k$ (al final de cada hora, día o periodo similar, no necesariamente de igual duración) de forma que los tiempos de vida de las unidades muestrales se registran agrupados en los (k+1) intervalos $(t_{i-1},t_i]$ $i=1,\ldots,k+1$.

Además, en estos instantes t_i, la variable se censura aleatoriamente según una distribución de probabilidad conocida (en cada t_i, se cambia o se deja de aplicar el tratamiento en un grupo de supervivientes, escogidos al azar con unas probabilidades prefijadas). Denotamos por C a la variable de censura. De esta forma, la información que recoge el investigador para cada individuo de la muestra es, o bien el intervalo $(t_{i-1},t_i]$ en el que éste muere (observación no censurada) o bien el instante t_i al que sobrevive (observación censurada).

Supondremos que un tiempo de vida censurado en el instante t_i es superior a dicho tiempo.

Definimos

$$p_i = P(T \in (t_{i-1},t_i]) = \int_{t_{i-1}}^{t_i} dF(t) \qquad i=1,\ldots,k+1$$

y

$$c_i = P(C = t_i) \qquad i=1,\dots,k$$

Denotaremos por $c_{k+1} = 1- \sum_{i=1}^{k} c_i$

Cada p_i representa la probabilidad de morir en $(t_{i-1},t_i]$; cada c_i la probabilidad de censura en t_i y c_{k+1} la probabilidad de no censurar. Supondremos que las c_i son conocidas y que ambas variables son independientes.

Estamos pues,ante un modelo de supervivencia censurado por la derecha, con la variable de censura discreta y la variable tiempo de vida discretizada por restricciones de observación.

La variable observada para cada individuo de la muestra, que denotamos por d_c, podrá presentar 2k+1 modalidades excluyentes (morir en cualquiera de los (k+1) intervalos de tiempo o censurar en cualquiera de los k instantes de tiempo) con lo que podremos expresarla como una variable discreta 2k-dimensional con distribución multinomial de parámetros 1 y vector de probabilidades,aquél cuyas componentes son las probabilidades correspondientes a 2k de las modalidades.

En caso de ausencia de censura la variable observada d_o presentará k+1 modalidades excluyentes (morir en cada intervalo $(t_{i-1},t_i]$ $i=1,\dots,k+1$) y podrá expresarse por una variable discreta k-dimensional con distribución multinomial de parámetros 1 y vector de probabilidades p con componentes p_i, $i=1,\dots,k$.

Tenemos así, la base intuitiva que nos lleva a considerar los experimentos ε_c (censurado) y ε_o (no censurado) que formalizamos en la sección 3.2.

3.2. ESTRUCTURA FORMAL

Desde un punto de vista formal, trabajaremos con una estructura matemática que consta de los siguientes elementos:

1.- Un espacio paramétrico $\mathcal{P}=\{p=(p_1,\dots,p_k)\in\mathbb{R}^k : p_i>0\ (i=1,\dots,k)$ y $\sum_{i=1}^{k} p_i<1\}$

2.- Una familia de experimentos ε_c, con $c=(c_1,\dots,c_k)$, $c_i\geqslant 0$ $(i=1,\dots,k)$, $\sum_{i=1}^{k} c_i\leqslant 1$. ε_c consiste en una observación de una variable 2k-dimensional (*) $d_c=(X_1,X_2,\dots,X_k,Y_1,Y_2,\dots,Y_k)$ cuya distribución para $p\in\mathcal{P}$ fijo, es multinomial de parámetros 1 y vector de probabilidades $p'=g(p)\in\mathbb{R}^{2k}$ definido como sigue:

$$p'=(p_1\sum_{j=1}^{k+1} c_j, p_2\sum_{j=2}^{k+1} c_j,\dots,p_k\sum_{j=k}^{k+1} c_j, c_1\sum_{j=2}^{k+1} p_j, c_2\sum_{j=3}^{k+1} p_j,\dots$$

$$\dots,c_k p_{k+1})\quad \text{siendo } c_{k+1}=1-\sum_{i=1}^{k} c_i \quad \text{y} \quad p_{k+1}=1-\sum_{i=1}^{k} p_i$$

Denotaremos por ε_o al experimento ε_c cuando $c=(0,0,\dots,0)$.

En este caso, las variables $Y_1,Y_2,\dots,Y_k$ del vector d_c resultan degeneradas en 0 y la variable observada $d_o=(X_1,X_2,\dots,X_k)$ sigue una distribución, para $p\in\mathcal{P}$ fijo, multinomial de parámetros 1 y p.

(*) Se sobreentiende, que si r de las k componentes de c son nulas, $r\leqslant k$, la variable observada d_c será una multinomial (2k-r)-dimensional.

Simplificaremos la notación, escribiendo $p_{(i)} = \sum_{j=i}^{k+1} p_j$ $(i=1,\ldots,k+1)$ y de manera similar $c_{(i)} = \sum_{j=i}^{k+1} c_j$ $(i=1,\ldots,k+1)$.

Denotaremos M_u, B_i para referirnos a las distribuciones multinomial y binomial respectivamente.

Dado que la variable d_c se distribuye $M_u(1,p')$, entonces las variables marginales $X_i \sim B_i(1,p_i c_{(i)})$ $i=1,\ldots,k$ e $Y_j \sim B_i(1,c_j p_{(j+1)})$ $j=1,\ldots,k$.

Definimos otro vector de parámetros $q=(q_1,\ldots,q_k)$ estrechamente relacionado con p y que nos será muy útil posteriormente

$$q_i = P(T \leq t_i | T > t_{i-1}) = \frac{p_i}{p_{(i)}} \qquad i=1,\ldots,k$$

$$q_{k+1} = P(T < t_{k+1} | T > t_k) = 1$$

Es fácil comprobar que entre p y q existe una correspondencia biyectiva. En particular, dado q, podemos obtener el vector de probabilidades p mediante las siguientes relaciones

$$p_1 = q_1$$

$$p_i = q_i(1-q_{i-1})(1-q_{i-2})\ldots(1-q_1) \qquad i=2,\ldots,k+1 \qquad (3.2.1)$$

Esta reparametrización genera un nuevo espacio paramétrico $\mathcal{Q} = \{q=(q_1,\ldots,q_k) \in \mathbb{R}^k : 0<q_i<1 \quad (i=1,\ldots,k)\}$ y consiguientemente las variables d_c y d_o, cuyos modelos probabilísticos dependen del parámetro p, $p \in \mathcal{P}$, pueden formalizarse mediante otros modelos que dependen del parámetro q, $q \in \mathcal{Q}$. En concreto las funciones de probabilidad de la variable d_c, respecto de p y q son respectivamente

$$f_c(x_1,x_2,\ldots,x_k,y_1,y_2,\ldots,y_k;p) = \prod_{i=1}^{k} (p_i c_{(i)})^{x_i} (c_i p_{(i+1)})^{y_i} (p_{k+1} c_{K+1})^{x_{k+1}}$$

$$\left\{ \begin{array}{l} x_i, y_i = 0,1 \\ \sum_{i=1}^{k} (x_i+y_i) \leqslant 1 \\ x_{k+1} = 1- \sum_{i=1}^{k} (x_i+y_i) \end{array} \right. \qquad (3.2.2)$$

y

$$f_c(x_1,x_2,\ldots,x_k,y_1,y_2,\ldots,y_k;q) = \prod_{i=1}^{k} q_i^{x_i} (1-q_i)^{x_{(i+1)}+y_{(i)}} c_{(i)}^{x_i} c_i^{y_i} c_{k+1}^{x_{k+1}} \qquad (3.2.3)$$

obtenida de la anterior a partir de las ecuaciones (3.2.1), y denotando como antes $x_{(i)} = \sum_{j=i}^{k+1} x_j$ $(i=1,\ldots,k+1)$ e $y_{(i)} = \sum_{j=i}^{k} y_j$ $(i=1,\ldots,k)$.

Análogamente, las funciones de probabilidad de la variable d_o respecto de p y q son respectivamente

$$f_o(x_1,\ldots,x_k;p) = \prod_{i=1}^{k+1} p_i^{x_i}$$

y

$$f_o(x_1,\ldots,x_k;q) = \prod_{i=1}^{k} q_i^{x_i} (1-q_i)^{x_{(i+1)}} \qquad \left\{ \begin{array}{l} x_i = 0,1 \\ \sum_{i=1}^{k} x_i \leqslant 1 \\ x_{k+1} = 1- \sum_{i=1}^{k} x_i \end{array} \right.$$

que son, como era de esperar, los valores particulares de las funciones f_c anteriores para $c=(0,\ldots,0)$ (Por convenio, tomamos $0^0=1$).

3.3. RELACION ENTRE LOS EXPERIMENTOS ε_c Y ε_o

Vamos a dar una serie de definiciones y resultados previos acerca de la comparación de experimentos que usaremos posteriormente para los experimentos ε_c y ε_o.

Sean $\varepsilon_X=\{(X,\Omega_X,\mathcal{A});P_\theta:\theta\in\Theta\}$ y $\varepsilon_Y=\{(Y,\Omega_Y,\mathcal{B});Q_\theta:\theta\in\Theta\}$ dos experimentos con el mismo espacio paramétrico.

Definición 3.3.1

Una transformación estocástica de ε_X a ε_Y es una función no negativa $\Pi(\cdot|\cdot)$ definida sobre $\mathcal{B}\times\Omega_X$ tal que:

1) Para cada $x\in\Omega_X$, $\Pi(\cdot|x)$ es una medida de probabilidad sobre $\mathcal{B}$
2) Para cada $B\in\mathcal{B}$, $\Pi(B|\cdot)$ es una función $\mathcal{A}$-medible sobre Ω_X.

Basándose en esta definición, Blackwell (1951, 1953) establece el concepto de suficiencia del experimento ε_X respecto del experimento ε_Y, que exponemos a continuación.

Definición 3.3.2

El experimento ε_X es suficiente para el experimento ε_Y y denotamos $\varepsilon_X\gtrsim\varepsilon_Y$ si existe una transformación estocástica $\Pi(\cdot|\cdot)$ de ε_X a ε_Y tal que:

$$Q_\theta(B) = \int_{\Omega_X} \Pi(B|x)\,dP_\theta(x)$$

para todo $B\in\mathcal{B}$ y $\theta\in\Theta$

El significado de esta definición es que si $\varepsilon_X\gtrsim\varepsilon_Y$ entonces es posible, por medio de una observación de X y una aleatorización auxiliar especificada por la distribución de probabilidad $\Pi(\cdot|x)$, generar una variable aleatoria $Z(X)$ que tenga la misma distribución que Y para todos los valores de θ.

Lehmann (1986; sec. 3.4) establece la definición anterior en la siguiente forma equivalente.

Definición 3.3.3

El experimento ε_X es suficiente para el experimento ε_Y si existe una función medible $h(X,U)$, donde U es una variable aleatoria independiente de X y de distribución conocida, tal que las distribuciones de $h(X,U)$ e Y son idénticas para todo $\theta \in \Theta$.

Incluimos, seguidamente, un resultado debido a Goel (1987) que establece, para modelos de supervivencia paramétricos, la suficiencia de un experimento no censurado respecto al correspondiente censurado aleatoriamente por la derecha (ver sección 2.1).

Teorema 3.3.1

Sean ε y ζ dos experimentos censurados aleatoriamente por la derecha basados en las variables (T,C_1) y (T,C_2) respectivamente, tal que $T \sim F_T(t,\theta)$ $\theta \in \Theta$ y $C_1 \overset{st}{\leq} C_2$. Entonces el experimento ζ es suficiente para el experimento ε.

Tomando $C_2=\infty$ en el teorema precedente, obtenemos el siguiente corolario.

Corolario 3.3.2

Un experimento no censurado basado en la variable aleatoria T es suficiente para el experimento censurado aleatoriamente por la derecha, basado en la variable aleatoria (T,C) para cualquier variable de censura C.

Con las definiciones y resultados precedentes, estamos ya en condiciones de comparar los experimentos ε_c y ε_o definidos en la sección 3.2.

Teorema 3.3.3

ε_o es suficiente para ε_c, para toda c.

Demostración

A primera vista parece que el resultado es consecuencia inmediata del corolario 3.3.2, sin embargo, los experimentos ε_o y ε_c no se corresponden formalmente con los experimentos no censurado y censurado respectivamente de un modelo de supervivencia, que aparecen en dicho corolario y cuya formalización, que exponíamos en el capítulo II (sección 2.1), recordamos a continuación:

Si llamamos T y C a las variables tiempo de vida y tiempo de censura respectivamente, el experimento no censurado consiste en la observación de T, mientras el experimento censurado - (aleatoriamente por la derecha) observa la variable (X,δ) definida por

$X = \min(T,C)$ y $\delta = I_{[T\leq C]}$, con $I_{[A]}$ la función indicador de A

Es claro que las variables d_o y d_c no pueden formalizarse de esta forma, sin embargo es sencillo encontrar una correspondencia biyectiva entre estas variables d_o y d_c y las correspondientes T y (X,δ) de un modelo de supervivencia censurado en el que ambas variables T y C sean discretas. Veámoslo.

Sean $p_i = P(T=t_i)$ $i=1,\ldots,k+1$, con $t_k < t_{k+1} < \infty$

y $\quad c_i = P(C=t_i) \quad i=1,\dots,k$

Establecemos la biyección $\varphi : d_o \longrightarrow T$ definida por

$$(0,0,\dots,\overset{i}{1},\dots,\overset{k}{0}) \longrightarrow t_i \qquad i=1,\dots,k$$

$$(0,\dots\dots\dots,0) \longrightarrow t_{k+1}$$

y la biyección $\psi : d_c \longrightarrow (X,\delta)$ definida por

$$(\overset{1}{0},\dots,\overset{i}{1},\dots,\overset{k}{0},\overset{1}{0},\dots,\overset{k}{0}) \longrightarrow (t_i,1) \qquad i=1,\dots,k$$

$$(\overset{1}{0},\dots\dots\dots,\overset{k}{0},\overset{1}{0},\dots,\overset{j}{1},\dots,\overset{k}{0}) \longrightarrow (t_j,0) \qquad j=1,\dots,k$$

$$(0,\dots\dots\dots,0,0,\dots,0) \longrightarrow (t_{k+1},1)$$

De esta forma, las variables T y (X,δ) son dos estadísticos suficientes para d_o y d_c respectivamente, por lo que la suficiencia del experimento que observa T respecto del experimento que observa (X,δ) (corolario 3.3.2) es equivalente a la de ε_o respecto de ε_c (DeGroot; 1966, sec. 3).

$\square$

Vamos a incluir por su interés y sencillez, una segunda demostración de este teorema que no requiere el resultado general del corolario 3.3.2. Consiste en comprobar directamente la definición de suficiencia de Blackwell; probaremos que $\varepsilon_o \gtrsim \varepsilon_c$ según la definición 3.3.3.

<u>Demostración</u>

Por definición, la variable d_o toma los valores

$d_o^i = (\overset{1}{0},\dots,\overset{i}{1},\dots,\overset{k}{0})$ con probabilidad p_i $(i=1,\dots,k)$ y

$d_o^{k+1} = (0,\dots\dots\dots,0)$ con probabilidad p_{k+1}

Sea U la variable aleatoria, independiente de d_o que toma los valores

$$u_j = (\overset{1}{0},\ldots,\overset{j}{1},\ldots,\overset{k}{0}) \quad \text{con probabilidad } c_j \ (j=1,\ldots,k) \text{ y}$$

$$u_{k+1} = (0,\ldots\ldots\ldots,0) \quad \text{con probabilidad } c_{k+1}$$

Definimos la función $h(d_o,U)$ de la forma siguiente

$$h(d_o^i,u_j) = \begin{cases} (d_o^i,u_{k+1}) & \text{si } i \leq j \\ (d_o^{k+1},u_j) & \text{si } i > j \end{cases}$$

Para todo $p \in P$ fijo, $h(d_o,U)$ tiene la misma distribución que d_c.

□

Una vez comparados los experimentos ε_c y ε_o, dedicamos las próximas secciones al estudio de las medidas paramétricas de información contenidas en dichos experimentos.

Dado que nuestro parámetro de interés p o alternativamente q, es k-variante, utilizaremos como medidas paramétricas de información adecuadas, la matriz de información de Fisher y las medidas de información reales basadas en dicha matriz (sección 1.3).

3.4. MATRIZ DE INFORMACION DE FISHER

Denotamos por $I_c(p)$ e $I_c(q)$ a las matrices de información de Fisher obtenidas a partir de los datos d_c, acerca de p y q respectivamente. Análogamente $I_o(p)$ e $I_o(q)$ para los datos d_o.

<u>Resultado 3.4.1</u>

La matriz simétrica de orden k, $I_c(p)$ tiene por elementos $I_c^{ij}(p)$ las siguientes funciones de p

$$I_c^{ii}(p) = \frac{c_{(i)}}{p_i} + \frac{c_{k+1}}{p_{k+1}} + \sum_{\ell=i}^{k} \frac{c_\ell}{p_{(\ell+1)}} \qquad i=1,\dots,k$$

$$I_c^{ij}(p) = \frac{c_{k+1}}{p_{k+1}} + \sum_{\ell=i}^{k} \frac{c_\ell}{p_{(\ell+1)}} \qquad j<i \ , \ i=2,\dots,k$$

Demostración

Es sencillo comprobar que $f_c(x_1,\dots,y_k;p)$ verifica las condiciones A,B,C,D de regularidad de la matriz de Fisher sustituyendo el símbolo integral por el sumatorio (sección 1.2), lo que permite calcular la casilla (i,j) de dicha matriz a partir de la expresión

$$I_c^{ij}(p) = -E_p\left[\frac{\partial^2}{\partial p_i \partial p_j} \log f_c(d_c;p)\right] \qquad i,j=1,\dots,k$$

De (3.2.2) tenemos

$$\log f_c(d_c;p) = \sum_{i=1}^{k} (x_i \log p_i + y_i \log p_{(i+1)}) + x_{k+1}\log p_{k+1} + H$$

siendo H una función que no depende de p.

Teniendo en cuenta que $p_{k+1} = (1-p_1-\dots-p_k)$

$$\frac{\partial}{\partial p_i} \log f_c = \frac{x_i}{p_i} - \sum_{\ell=i}^{k} \frac{y_\ell}{p_{(\ell+1)}} - \frac{x_{k+1}}{p_{k+1}} \qquad i=1,\dots,k$$

$$\frac{\partial^2}{\partial p_i^2} \log f_c = -\frac{x_i}{p_i^2} - \sum_{\ell=i}^{k} \frac{y_\ell}{p_{(\ell+1)}^2} - \frac{x_{k+1}}{p_{k+1}^2} \qquad i=1,\dots,k$$

y

$$\frac{\partial^2}{\partial p_i \partial p_j} \log f_c = -\sum_{\ell=i}^{k} \frac{y_\ell}{p_{(\ell+1)}^2} - \frac{x_{k+1}}{p_{k+1}^2} \qquad i=1,\dots,k \ ; \ j<i$$

Tomando esperanzas (sabemos que $X_i \sim B_i(1,p_i c_{(i)})$ e $Y_j \sim B_i(1,c_j p_{(j+1)})$ $i,j=1,\dots,k$) y cambiando de signo

$$I_c^{ii}(p) = \frac{p_i c_{(i)}}{p_i^2} + \sum_{\ell=i}^{k} \frac{c_\ell p_{(\ell+1)}}{p_{(\ell+1)}^2} + \frac{p_{k+1}c_{k+1}}{p_{k+1}^2} \qquad i=1,\dots,k$$

e

$$I_c^{ij}(p) = \sum_{\ell=i}^{k} \frac{c_\ell p_{(\ell+1)}}{p^2_{(\ell+1)}} + \frac{p_{k+1} c_{k+1}}{p^2_{k+1}} \qquad i=2,\dots,k \ ; \ j<i$$

□

Observaciones

1.- Otra forma alternativa de obtener la matriz $I_c(p)$ sería partiendo de la expresión de la matriz de Fisher para datos agrupados (ver sección 1.1).

Si denotamos por p'_ℓ $\ell=1,\dots,2k$, a las componentes del vector p', con $p'_{2k+1}=1-\sum_{\ell=1}^{2k} p'_\ell$, entonces

$$I_c^{ij}(p) = \sum_{\ell=1}^{2k+1} \left(\frac{\partial}{\partial p_i} \log p'_\ell\right)\left(\frac{\partial}{\partial p_j} \log p'_\ell\right) p'_\ell \qquad i,j=1,\dots,k$$

que podemos expresar de manera más explícita como sigue

$$\sum_{\ell=1}^{k} \left(\frac{\partial}{\partial p_i} \log p_\ell c_{(\ell)}\right)\left(\frac{\partial}{\partial p_j} \log p_\ell c_{(\ell)}\right) p_\ell c_{(\ell)} +$$

$$\sum_{\ell=1}^{k} \left(\frac{\partial}{\partial p_i} \log c_\ell p_{(\ell+1)}\right)\left(\frac{\partial}{\partial p_j} \log c_\ell p_{(\ell+1)}\right) c_\ell p_{(\ell+1)} +$$

$$+ \left(\frac{\partial}{\partial p_i} \log c_{k+1} p_{k+1}\right)\left(\frac{\partial}{\partial p_j} \log c_{k+1} p_{k+1}\right) c_{k+1} p_{k+1}$$

Operaciones de cálculo muy sencillas nos conducen al resultado 3.4.1.

2.- Llamamos la atención sobre la forma peculiar de la matriz $I_c(p)$; en cualquiera de las filas o columnas de $I_c(p)$ todos los elementos anteriores a la diagonal son iguales; es decir

$$I_c^{ij_1}(p) = I_c^{ij_2}(p) \quad \forall i, j_1, j_2 \text{ siendo } \begin{cases} j_1 < i \\ j_2 < i \end{cases}$$

Corolario 3.4.2

La matriz $I_o(p)$ tiene por elementos $I_o^{ij}(p)$ las siguientes funciones de p

$$I_o^{ii}(p) = \frac{1}{p_i} + \frac{1}{p_{k+1}} \qquad i=1,\dots,k$$

$$I_o^{ij}(p) = \frac{1}{p_{k+1}} \qquad i \neq j \quad i,j=1,\dots,k$$

Demostración

Teniendo en cuenta que la variable d_o representa la variable d_c con $c=(0,0,\dots,0)$, basta sustituir estas componentes del vector c en el resultado 3.4.1, para completar la demostración.

□

Observación

Nótese que obtendríamos la misma matriz de información del corolario anterior, si el vector de probabilidades $c = (0,0,\dots,0,c_k)$ con $c_k > 0$.

Este resultado era de esperar intuitivamente, ya que a efectos de información acerca de $p=(p_1,p_2,\dots,p_k)$, es igual censurar sólo en t_k que no censurar en ningún instante.

La explicación formal de este hecho se apoya en el siguiente resultado.

Resultado 3.4.3

Sea la variable aleatoria d_{c^*} con $c^*=(0,\dots,0,c_k)$ y $c_k > 0$, con función de probabilidad

$$f_{c*}(x_1,x_2,\dots,x_k,y_k;p) = \prod_{i=1}^{k} p_i^{x_i}(c_k p_{k+1})^{y_k}(c_{k+1}p_{k+1})^{x_{k+1}}$$

$$x_{k+1} = 1-\sum_{i=1}^{k} x_i - y_k$$

Entonces, el estadístico $T(x_1,x_2,\dots,x_k,y_k) = d_o =$ $= (x_1,x_2,\dots,x_k)$ es suficiente (conjuntamente) para p.

Demostración

$$f_{c*}(x_1,\dots,x_k,y_k;p) = \prod_{i=1}^{k} p_i^{x_i}\, p_{k+1}^{1-\sum_{i=1}^{k} x_i}\, c_k^{y_k}\, c_{k+1}^{x_{k+1}} =$$

$$= f_o(d_o;p)\,h(x_1,\dots,x_k,y_k)$$

El resultado es consecuencia del teorema de factorización.

$\square$

A partir de este resultado, la igualdad de las matrices $I_o(p)$ e $I_{c*}(p)$ se deduciría formalmente de la propiedad (3) de la matriz de Fisher (sección 1.2).

Resultado 3.4.4

$$I_o(p) \geq I_c(p) \quad \text{para toda } c \text{ y } p \in P$$

Demostración

Vimos, en el teorema 3.3.3 que

$$\varepsilon_o \gtrsim \varepsilon_c \quad \text{para toda } c$$

La propiedad (5) de suficiencia de experimentos de la matriz de Fisher (sección 1.2) conduce al resultado.

$\square$

Corolario 3.4.5

$I_o^{ii}(p) \geq I_c^{ii}(p)$ para toda c, i=1,...,k y $p \in P$

Demostración

Si una matriz es semidefinida positiva, los elementos de su diagonal principal son no negativos.

Por el resultado 3.4.4 tenemos que $I_o(p)-I_c(p) \geq 0$ para toda c y $p \in P$ por lo que el resultado se deduce inmediatamente.

$\square$

Si tenemos en cuenta que $I_o^{ii}(p)$ e $I_c^{ii}(p)$ representan las cantidades de información de Fisher (unidimensionales) proporcionadas por ε_o y ε_c respectivamente, sobre p_i (dados los valores del resto de parámetros perturbadores p_j $j\neq i$) y denotamos consecuentemente a dichas medidas de información por $I_o(p_i)$ e $I_c(p_i)$ respectivamente, el corolario 3.4.5 expresa un resultado esperado intuitivamente, como es la pérdida de información acerca de cada parámetro p_i, a causa de la censura, es decir:

$$I_o(p_i) \geq I_c(p_i) \qquad \text{para toda c, i=1,...,k y } p \in P$$

Resultado 3.4.6

La matriz $I_c(q)$ es diagonal para toda c y $q \in Q$, con elementos diagonales $I_c^{ii}(q)$ las siguientes funciones de q

$$I_c^{11}(q) = \frac{1}{q_1(1-q_1)}$$

$$I_c^{ii}(q) = \frac{c_{(i)}(1-q_{i-1})(1-q_{i-2})\dots(1-q_1)}{q_i(1-q_i)} \qquad i=2,\dots,k$$

Demostración

Al igual que en el resultado 3.4.1 se cumplen las condiciones de regularidad que permiten obtener el elemento (i,j) de la matriz $I_c(q)$ de la siguiente forma

$$I_c^{ij}(q) = -E_q \left[\frac{\partial^2}{\partial q_i \partial q_j} \log f_c(d_c;q) \right] \qquad i,j=1,2,\dots,k$$

De (3.2.3) tenemos

$$\log f_c(d_c;q) = \sum_{i=1}^{k} \left[x_i \log q_i + (x_{(i+1)}+y_{(i)}) \log(1-q_i) \right] + H$$

Derivando respecto a q_i

$$\frac{\partial}{\partial q_i} \log f_c = \frac{x_i}{q_i} - \frac{(x_{(i+1)}+y_{(i)})}{1-q_i} \qquad i=1,\dots,k$$

$$\frac{\partial^2}{\partial q_i^2} \log f_c = -\frac{x_i}{q_i^2} - \frac{(x_{(i+1)}+y_{(i)})}{(1-q_i)^2} \qquad i=1,\dots,k$$

y

$$\frac{\partial^2}{\partial q_i \partial q_j} \log f_c = 0 \qquad i\neq j \quad i,j=1,\dots,k$$

Dado que $X_{(i+1)}+Y_{(i)} \sim B_i(1,c_{(i)}p_{(i+1)})$ $i=1,\dots,k$ (Wilks; 1962, prob. 6.12 y 6.14 p. 151) tomando esperanzas y cambiando de signo, tenemos:

$$I_c^{ii}(q) = \frac{p_i c_{(i)}}{q_i^2} + \frac{c_{(i)} p_{(i+1)}}{(1-q_i)^2} \qquad i=1,\dots,k$$

utilizando las ecuaciones (3.2.1) tenemos

$$I_c^{ii}(q) = c_{(i)} \left[\frac{q_i(1-q_{i-1})(1-q_{i-2})\dots(1-q_1)}{q_i^2} + \right.$$

$$\left. + \frac{(1-q_i)\dots(1-q_1)}{(1-q_i)^2} \right] = c_{(i)} \left[(1-q_{i-1})\dots(1-q_1)\left(\frac{1}{q_i} + \frac{1}{1-q_i}\right) \right] =$$

$$= \frac{c_{(i)}(1-q_{i-1})\dots(1-q_1)}{q_i(1-q_i)} \qquad i=2,\dots,k$$

Claramente, para i=1

$$I_c^{11}(q) = \frac{p_1}{q_1^2} + \frac{p_{(2)}}{(1-q_1)^2} = \frac{q_1}{q_1^2} + \frac{(1-q_1)}{(1-q_1)^2} = \frac{1}{q_1(1-q_1)}$$

$\square$

Corolario 3.4.7

La matriz $I_o(q)$ es diagonal para todo $q \in Q$ siendo sus elementos diagonales $I_o^{ii}(q)$ las siguientes funciones de q

$$I_o^{11}(q) = \frac{1}{q_1(1-q_1)}$$

$$I_o^{ii}(q) = \frac{(1-q_{i-1})(1-q_{i-2})\dots(1-q_1)}{q_i(1-q_i)} \qquad i=2,\dots,k$$

Demostración

Es consecuencia inmediata del resultado 3.4.6. Ahora c = = (0,0,...,0) por lo que $c_{(i)}=1$ $i=1,\dots,k$.

$\square$

Al igual que ocurría respecto de p (observación del corolario 3.4.2), si la variable de censura tiene por vector de probabilidad $c^*=(0,0,\dots,c_k)$, $c_k > 0$, a efectos de información equivale a no censurar las observaciones ($c=(0,\dots,0)$), es decir $I_{c^*}(q) = I_o(q)$ para todo $q \in Q$.

Resultado 3.4.8

$I_o(q) \geqslant I_c(q)$ para toda c y $q \in Q$

Demostración

Sabemos que $\varepsilon_o \gtrsim \varepsilon_c$ para toda c cuando el espacio paramétrico es P.

Si bien el concepto de suficiencia de experimentos se establece en relación a un espacio paramétrico determinado (definición 3.3.2), es sencillo comprobar que dicho concepto permanece inalterable para cualquier reparametrización biyectiva del espacio paramétrico original.

En particular, podemos escribir $\varepsilon_o \gtrsim \varepsilon_c$ para toda c cuando el espacio paramétrico considerado es Q. La demostración se concluye de forma similar a la del resultado 3.4.4.

$\square$

<u>Corolario 3.4.9</u>

$$I_o^{ii}(q) \geq I_c^{ii}(q) \quad \text{para toda } c, \quad i=1,\dots,k \quad y \quad q \in Q$$

<u>Demostración</u>

Similar a la del Corolario 3.4.5.

$\square$

Dado que $I_c^{ii}(q)$ e $I_o^{ii}(q)$ representan las cantidades de información de Fisher proporcionadas por ε_c y ε_o respectivamente, sobre q_i (dados los valores del resto de parámetros perturbadores q_j $j \neq i$), el corolario 3.4.9 expresa la pérdida de información sobre cada q_i, a causa de la censura, que podemos escribir en una notación más adecuada como sigue

$$I_o(q_i) \geq I_c(q_i) \quad \text{para toda } c, \; i=1,\dots,k \quad y \quad q \in Q$$

Obviamente $I_o(q_1) = I_c(q_1)$ para toda c.

3.5. RELACION ENTRE LAS MATRICES DE INFORMACION $I_c(p)$ e $I_c(q)$

El siguiente resultado indica la expresión que relaciona las matrices $I_c(p)$ e $I_c(q)$ y por tanto la forma de obtener la -

primera a partir de la segunda (más fácil de calcular).

Resultado 3.5.1

$I_c(p) = A'I_c(q)A$ para toda c, siendo A la matriz triangular

$$\begin{pmatrix} 1 & 0 & 0 & 0 \cdots\cdots 0 \\ \dfrac{p_2}{p_{(2)}^2} & \dfrac{1}{p_{(2)}} & 0 & 0 \cdots\cdots 0 \\ \dfrac{p_3}{p_{(3)}^2} & \dfrac{p_3}{p_{(3)}^2} & \dfrac{1}{p_{(3)}} & 0 \cdots\cdots 0 \\ \vdots & \vdots & \vdots & \\ \dfrac{p_k}{p_{(k)}^2} & \dfrac{p_k}{p_{(k)}^2} & \dfrac{p_k}{p_{(k)}^2} \cdots\cdots & \dfrac{1}{p_{(k)}} \end{pmatrix}$$

Demostración

Entre p y q existe una correspondencia biyectiva, de forma que

$$q = (q_1,\ldots,q_k) = f(p) = (p_1, \frac{p_2}{p_{(2)}}, \ldots, \frac{p_k}{p_{(k)}})$$

Entonces $I_c(p) = A'I_c(q)A$ siendo A la matriz cuadrada cuyo elemento (i,j) es $a_{ij}=\frac{\partial q_i}{\partial p_j}$, i,j=1,...,k (ver (1.3.8), apdo.1.3.4)

A partir de aquí el resultado es inmediato.

□

Otra expresión de la relación entre $I_c(p)$ e $I_c(q)$, que utilizaremos posteriormente, aparece en el siguiente corolario.

Corolario 3.5.2

$$I_c(p) = I_c^{11}(q)\begin{pmatrix}1\\0\\0\\ \vdots\\0\end{pmatrix}(1\ 0\ 0\ldots 0)+I_c^{22}(q)\begin{pmatrix}p_2/p_{(2)}^2\\1/p_{(2)}\\0\\ \vdots\\0\end{pmatrix}\left(\frac{p_2}{p_{(2)}^2}\ \frac{1}{p_{(2)}}\ 0\ldots 0\right) +$$

$$+\ldots+I_c^{kk}(q)\begin{pmatrix}p_k/p_{(k)}^2\\p_k/p_{(k)}^2\\ \vdots\\1/p_{(k)}\end{pmatrix}\left(\frac{p_k}{p_{(k)}^2}\ \frac{p_k}{p_{(k)}^2}\ \ldots\ \frac{1}{p_{(k)}}\right)$$

Demostración

Es consecuencia del resultado 3.5.1, por ser $I_c(q)$ una matriz diagonal.

□

Corolario 3.5.3

$I_o(p) = A'I_o(q)A$ siendo A la matriz definida en el resultado 3.5.1.

Demostración

Basta tomar c=(0,0,...,0) en el resultado 3.5.1.

□

Observación

La matriz $I_c(q)$ se obtendría a partir de $I_c(p)$ mediante la expresión $I_c(q) = B'I_c(p)B$ siendo B la matriz cuadrada de elemento (i,j)

$$b_{ij} = \frac{\partial p_i}{\partial q_j} \qquad i,j=1,\dots,k \qquad \text{(ver (1.3.9),apdo.1.3.4)}$$

Además $|A| = \prod_{i=1}^{k} \frac{1}{P_{(i)}} > 0$, por lo que $B=A^{-1}$

3.6. MEDIDAS DE INFORMACION REALES

Dedicamos esta sección al cálculo de las medidas de información paramétricas reales, basadas en la matriz de Fisher, que definíamos en la sección 1.3. Siguiendo la notación de dicha - sección, escribiremos $S_c(p)$ y $S_c(q)$ para referirnos a la medida de información paramétrica definida en el epígrafe 1.3.3.1, basada en las matrices $I_c(p)$ e $I_c(q)$ respectivamente. Análogamente $D_c(p)$ y $D_c(q)$ denotarán la medida de información paramétrica definida en el epígrafe 1.3.3.2, basada en las matrices $I_c(p)$ e $I_c(q)$ respectivamente. Por último $M_c(p)$ y $M_c(q)$ representan la medida de información paramétrica definida en el epígrafe 1.3.3.3 basada en las matrices correspondientes.

El subíndice o en lugar del c en todas las medidas de información anteriores indicará que el vector $c=(0,\dots,0)$.

3.6.1. MEDIDA DE INFORMACION S_c

Por definición $\quad S_c(p) = \alpha\left[\text{Traza}(I_c(p))\right] = \alpha \sum_{i=1}^{k} I_c^{ii}(p)$

donde α es una constante positiva prefijada.

Análogamente $\quad S_c(q) = \alpha \sum_{i=1}^{k} I_c^{ii}(q)$

Resultado 3.6.1

$$S_c(p) = \alpha\left[\sum_{i=1}^{k} c_i\left(\frac{1}{p_1}+\frac{1}{p_2}+\ldots+\frac{1}{p_i}+\frac{i}{p_{(i+1)}}\right)+c_{k+1}\left(\frac{1}{p_1}+\frac{1}{p_2}+\ldots+\frac{1}{p_k}+\frac{k}{p_{k+1}}\right)\right]$$

Demostración

Observando el resultado 3.4.1.

$$\text{Traza}\left[I_c(p)\right] = \sum_{i=1}^{k}\frac{c_{(i)}}{p_i}+\sum_{i=1}^{k}\sum_{\ell=i}^{k}\frac{c_\ell}{p_{(\ell+1)}}+k\,\frac{c_{k+1}}{p_{k+1}} =$$

$$= \sum_{i=1}^{k}\left(\frac{c_i+c_{i+1}+\ldots+c_k+c_{k+1}}{p_i}+\sum_{\ell=i}^{k}\frac{c_\ell}{p_{(\ell+1)}}\right)+k\,\frac{c_{k+1}}{p_{k+1}}$$

Sacando factor común c_i dentro del paréntesis se obtiene el resultado.

□

Corolario 3.6.2

$$S_o(p) = \alpha\left[\sum_{i=1}^{k}\frac{1}{p_i}+\frac{k}{p_{k+1}}\right]$$

Demostración

Ahora $c_i=0$ $i=1,\ldots,k$, $c_{k+1}=1$. Sustituyendo estos valores en el resultado 3.6.1 se completa la demostración.

□

Resultado 3.6.3

$$S_c(q) = \alpha\left[\frac{1}{q_1(1-q_1)}+\sum_{i=2}^{k}\frac{c_{(i)}(1-q_{i-1})(1-q_{i-2})\ldots(1-q_1)}{q_i(1-q_i)}\right]$$

Demostración

Inmediata del resultado 3.4.6.

□

Corolario 3.6.4

$$S_o(q) = \alpha\left[\frac{1}{q_1(1-q_1)} + \sum_{i=2}^{k} \frac{(1-q_{i-1})\dots(1-q_1)}{q_i(1-q_i)}\right]$$

Demostración

Resulta inmediata del resultado 3.6.3, pues $c_{(i)}=1$, $i=1,\dots,k$

□

Resultado 3.6.5

$S_c(p) > S_c(q)$ para toda c

Demostración

Bastaría ver que cada elemento de la diagonal de $I_c(p)$ es mayor que el correspondiente de $I_c(q)$.

Por el corolario 3.5.2 sabemos que:

$$I_c^{ii}(p) = I_c^{ii}(q)\ \frac{1}{p^2_{(i)}} + \sum_{j=i+1}^{k} I_c^{jj}(q)\ \frac{p_j^2}{p^4_{(j)}} > I_c^{ii}(q)$$

$$i=1,\dots,k$$

Sumando en i y multiplicando por α en ambos miembros, obtenemos el resultado.

□

3.6.2. MEDIDA DE INFORMACION D_c

Por definición $D_c(p) = \left[|I_c(p)|\right]^{\alpha}$ donde α es una constante positiva prefijada.

Análogamente $D_c(q) = \left[|I_c(q)|\right]^{\alpha}$

Debido a su sencillez, calcularemos en primer lugar $D_c(q)$ y a partir de ésta $D_c(p)$.

Resultado 3.6.6

$$D_c(q) = \left[\frac{\prod\limits_{i=1}^{k-2}(1-q_i)^{k-1-i}}{1-q_k}\prod_{i=1}^{k}\frac{c_{(i)}}{q_i}\right]^{\alpha}$$

Demostración

Por ser $I_c(q)$ una matriz diagonal, su determinante es el producto de los elementos de dicha diagonal, es decir

$$|I_c(q)| = \prod_{i=1}^{k} I_c^{ii}(q)$$

A la vista del resultado 3.4.6.

$$\prod_{i=1}^{k} I_c^{ii}(q) = \frac{1}{q_1(1-q_1)}\prod_{i=2}^{k}\frac{c_{(i)}(1-q_{i-1})\dots(1-q_1)}{q_i(1-q_i)} =$$

$$= \prod_{i=2}^{k} c_{(i)}\,\frac{1}{q_1(1-q_1)}\,\frac{(1-q_1)}{q_2(1-q_2)}\cdot\ \dots\ \cdot\frac{(1-q_{k-1})(1-q_{k-2})\dots(1-q_1)}{q_k(1-q_k)} =$$

$$= \prod_{i=1}^{k}\frac{c_{(i)}}{q_i}\,\frac{\prod\limits_{i=1}^{k-2}(1-q_i)^{k-1-i}}{(1-q_k)}$$

$\square$

Corolario 3.6.7

$$D_o(q) = \left[\frac{1}{q_k q_{k-1}(1-q_k)} \prod_{i=1}^{k-2} \frac{(1-q_i)^{k-1-i}}{q_i}\right]^{\alpha}$$

Demostración

Inmediata del resultado 3.6.6.

□

Corolario 3.6.8

$D_c(q) = 0$ para todo $q \in Q \iff c_{(k)}=0$

Demostración

$\Rightarrow)$

Si $D_c(q) = 0 \Rightarrow c_{(i)}=0$ para algún $i=1,\dots,k \Rightarrow c_{(k)}=0$
(por ser $0 < q_i < 1 \quad \forall\ i=1,\dots,k$)

$\Leftarrow)$

Si $c_{(k)} = 0 \Rightarrow D_c(q) = 0$

□

Observación

La medida de información $D_c(q)$ se comporta como una buena medida de información (en relación a sus propiedades) excepto en el caso en que $c_{(k)}=0$. Sin embargo el resultado es explicable en términos de información, ya que $c_{(k)}=0$ indica que ninguna observación puede caer en el intervalo $(t_{k-1}, t_k]$ por lo que la información acerca de q_k es cero $(I_c^{kk}(q)=0)$. En las medidas $S_c(q)$ y $M_c(q)$ ésto se reflejaría en un cero, como último sumando; $D_c(q)$ se anula, siendo en este sentido una medida de infor mación más sensible que aquéllas.

Una explicación más amplia de este hecho aparece en el resumen del final del apartado 1.3.3. Allí proponemos una medida de información alternativa D_c^*, basada en D_c, para el caso de que la matriz de Fisher sea singular. En particular para la matriz $I_c(q)$, dicha medida será:

$$D_c^*(q) = \left[|I+I_c(q)|^{\alpha}\right]-1 = \left[\prod_{i=1}^{k}(1+I_c^{ii}(q))\right]^{\alpha}-1$$

siendo α una constante positiva prefijada.

Corolario 3.6.9

$I_c(q)$ es definida positiva para todo $q \in Q \Leftrightarrow c_{(k)} \neq 0$
$(c_k+c_{k+1} > 0)$

Demostración

Por la propiedad A.2 del Apéndice

$I_c(q)$ definida positiva $\Leftrightarrow |I_c(q)| \neq 0 \Leftrightarrow c_{(k)} \neq 0$

dándose la última equivalencia como consecuencia del corolario 3.6.8.

$\square$

Como consecuencia de este corolario,

Corolario 3.6.10

$I_o(q)$ es definida positiva para todo $q \in Q$

Demostración

En este caso $c_{(k)}=c_{k+1}=1$.

$\square$

Resultado 3.6.11

$$D_c(p) = \left[\frac{1}{p_{k+1}} \prod_{i=1}^{k} \frac{c_{(i)}}{p_i} \right]^{\alpha}$$

Demostración

La propiedad 4 de la medida D_c (epígrafe 1.3.4.2) establece que

$$D_c(p) = \left[|A|^2 \right]^{\alpha} D_c(q) \qquad (3.6.1)$$

siendo A la matriz definida en el resultado 3.5.1, cuyo determinante $|A| = \prod_{i=1}^{k} \frac{1}{p_{(i)}}$. Esto, junto con el resultado 3.6.6 hace que:

$$D_c(p) = \left[\prod_{i=1}^{k} \frac{1}{p^2_{(i)}} \right]^{\alpha} \left[\prod_{i=1}^{k-2} \frac{(1-q_i)^{k-1-i}}{1-q_k} \prod_{i=1}^{k} \frac{c_{(i)}}{q_i} \right]^{\alpha}$$

Sabemos que $q_i = \frac{p_i}{p_{(i)}}$ $i=1,\ldots,k$, por lo que:

$\prod_{j=1}^{i-1} (1-q_j) = p_{(i)}$ y $1-q_k = \frac{p_{k+1}}{p_{(k)}}$ (teniendo en cuenta las ecuaciones (3.2.1))

entonces:

$$\frac{\prod_{i=1}^{k-2} (1-q_i)^{k-1-i}}{1-q_k} = \prod_{i=2}^{k-1} p_{(i)} \frac{p_{(k)}}{p_{k+1}} = \frac{1}{p_{k+1}} \prod_{i=1}^{k} p_{(i)} \qquad (\text{ya que } p_{(1)}=1)$$

sustituyendo resulta que

$$D_c(p) = \left[\frac{1}{p_{k+1}} \prod_{i=1}^{k} \frac{1}{p^2_{(i)}} p_{(i)} \frac{c_{(i)}}{p_i} p_{(i)} \right]^{\alpha} = \left[\frac{1}{p_{k+1}} \prod_{i=1}^{k} \frac{c_{(i)}}{p_i} \right]^{\alpha}$$

□

Corolario 3.6.12

$$D_o(p) = \left[\prod_{i=1}^{k+1} \frac{1}{p_i}\right]^{\alpha}$$

Demostración

Inmediata del resultado 3.6.11. Basta tomar $c=(0,\ldots,0)$.

□

Corolario 3.6.13

$D_c(p) = 0$ para todo $p \in P \iff c_{(k)} = 0$

Demostración

$$|A| = \prod_{i=1}^{k} \frac{1}{p_{(i)}} > 0$$

El resultado es consecuencia del corolario 3.6.8 y de (3.6.1).

□

Corolario 3.6.14

$I_c(p)$ es definida positiva para todo $p \in P \iff c_{(k)} \neq 0$

Demostración

La matriz A es no singular, entonces

$I_c(p) = A'I_c(q)A$ es definida positiva $\iff$ $I_c(q)$ es definida positiva (Rao, 1973, pág. 35).

El resultado se concluye del corolario 3.6.9.

□

Como consecuencia inmediata

Corolario 3.6.15

$I_o(p)$ es definida positiva para todo $p \in P$

Demostración

Ahora $c_{(k)} = c_{k+1} = 1$ (no hay censura).

□

Resultado 3.6.16

$D_c(p) > D_c(q)$ para toda c

Demostración

Es consecuencia de la expresión (3.6.1) y del hecho de que

$$|A| = \prod_{i=1}^{k} \frac{1}{p_{(i)}} > 1$$

□

3.6.3. MEDIDA DE INFORMACION M_c

Por definición $M_c(p) = \|I_c(p)\| = \left[\sum_{i=1}^{k} \sum_{j=1}^{k} (I_c^{ij}(p))^2\right]^{1/2}$

Análogamente se define $M_c(q)$

Resultado 3.6.17

$$M_c(p) = \left[\sum_{i=1}^{k} \left(\frac{c_{(i)}}{p_i} + \frac{c_{k+1}}{p_{k+1}} + \sum_{\ell=i}^{k} \frac{c_\ell}{p_{(\ell+1)}}\right)^2 + \sum_{i=1}^{k} 2(i-1)\left(\frac{c_{k+1}}{p_{k+1}} + \sum_{\ell=i}^{k} \frac{c_\ell}{p_{(\ell+1)}}\right)^2\right]^{1/2}$$

Demostración

Debido a la simetría de $I_c(p)$ podemos escribir

$$\|I_c(p)\|^2 = \sum_{i=1}^{k} [I_c^{ii}(p)]^2 + \sum_{i=2}^{k} \sum_{j<i} 2[I_c^{ij}(p)]^2$$

A la vista de la forma peculiar de la matriz $I_c(p)$ (ver la 2ª observación del resultado 3.4.1) resulta que

$$\|I_c(p)\|^2 = \sum_{i=1}^{k} [I_c^{ii}(p)]^2 + \sum_{i=2}^{k} 2(i-1)[I_c^{i1}(p)]^2$$

El resultado 3.4.1 completa la demostración.

□

Corolario 3.6.18

$$M_o(p) = \left[\sum_{i=1}^{k} \frac{1}{p_i^2} + \frac{2}{p_{k+1}} \sum_{i=1}^{k} \frac{1}{p_i} + \frac{k^2}{p_{k+1}^2}\right]^{1/2}$$

Demostración

Tomando $c=(0,0,\dots,0)$ y sustituyendo en el resultado 3.6.17 obtenemos

$$\|I_o(p)\|^2 = \sum_{i=1}^{k}\left[\left(\frac{1}{p_i} + \frac{1}{p_{k+1}}\right)^2 + 2(i-1)\left(\frac{1}{p_{k+1}}\right)^2\right] =$$

$$= \sum_{i=1}^{k} \left(\frac{1}{p_i^2} + \frac{1}{p_{k+1}^2} + \frac{2}{p_i p_{k+1}} + (i-1)\frac{2}{p_{k+1}^2}\right) =$$

$$= \frac{1}{p_{k+1}^2} \sum_{i=1}^{k} (2i-1) + \frac{2}{p_{k+1}} \sum_{i=1}^{k} \frac{1}{p_i} + \sum_{i=1}^{k} \frac{1}{p_i^2} =$$

$$= \frac{k^2}{p_{k+1}^2} + \frac{2}{p_{k+1}} \sum_{i=1}^{k} \frac{1}{p_i} + \sum_{i=1}^{k} \frac{1}{p_i^2}$$

□

Resultado 3.6.19

$$M_c(q) = \left[\frac{1}{q_1^2(1-q_1)^2} + \sum_{i=2}^{k} \frac{c_{(i)}^2 (1-q_{i-1})^2 (1-q_{i-2})^2 \cdots (1-q_1)^2}{q_i^2(1-q_i)^2}\right]^{1/2}$$

Demostración

Por ser $I_c(q)$ diagonal para toda c

$\|I_c(q)\|^2 = \sum_{i=1}^{k} (I_c^{ii}(q))^2$. El resultado 3.4.6 completa la demostración.

□

Corolario 3.6.20

$$M_o(q) = \left[\frac{1}{q_1^2(1-q_1)^2} + \sum_{i=2}^{k} \frac{(1-q_{i-1})^2\ldots(1-q_1)^2}{q_i^2(1-q_i)^2}\right]^{1/2}$$

Demostración

Inmediata del resultado 3.6.19, ya que $c_{(i)}=1$ $i=1,\ldots,k$.

□

Resultado 3.6.21

$M_c(p) > M_c(q)$ para toda c

Demostración

Sabemos que

$I_c^{ii}(p) > I_c^{ii}(q) \geqslant 0$ $i=1,\ldots,k$ (demostración del resultado 3.6.5).

Entonces, elevando al cuadrado y sumando en i tenemos

$$\sum_{i=1}^{k} [I_c^{ii}(p)]^2 > \sum_{i=1}^{k} [I_c^{ii}(q)]^2 = \|I_c(q)\|^2$$

Pero, por definición

$$\|I_c(p)\|^2 \geqslant \sum_{i=1}^{k} [I_c^{ii}(p)]^2$$

a partir de aquí, el resultado se obtiene de forma inmediata.

□

3.7. PERDIDA DE INFORMACION. MEDIDAS MATRICIALES

Una vez calculadas las medidas de información adecuadas - acerca del parámetro p o q, nos interesa evaluar la pérdida de información que, acerca de dicho parámetro, se produce a causa de la censura, es decir, al observar ε_c en vez de ε_o. La forma de abordar este tema quedó expuesta con detalle en el capítulo II. Allí propusimos dos caminos, uno que podemos denominar "matricial", que abordaremos a lo largo de esta sección, adecuado cuando la medida de información utilizada es la matriz de Fisher; y otro que denominamos "real", que trataremos en la próxima sección, cuando las medidas de información son las medidas reales basadas en la matriz de Fisher.

El método matricial consiste en definir dos medidas matriciales de la pérdida de información, estrechamente relacionadas entre si, a saber, la matriz de pérdidas (apartado 2.3.1) y la matriz de eficiencias (apartado 2.4.2) y a partir de éstas definir medidas reales de la pérdida relativa de información (apartado 2.3.2) y de la eficiencia relativa (apartado - 2.4.3).

Siguiendo la notación del Capítulo II escribiremos $L_c(p)$ y $L_c(q)$ para referirnos a la matriz de pérdidas, al observar ε_c en vez de ε_o, acerca de p y q respectivamente. Denotaremos, asímismo por $R_c(p)$ y $R_c(q)$ a la matriz de eficiencias, de ε_c comparado con ε_o, acerca de p y q respectivamente. De igual for ma $L_c^i(p)$ $i=1,2,3$ denotarán las medidas reales de la pérdida relativa de información acerca de p, basadas en la matriz $L_c(p)$ y análogamente para q. Por último $R_c^i(p)$ $i=1,2,3$ expresarán - las medidas reales de la eficiencia relativa acerca de p, basadas en la matriz $R_c(p)$ y similarmente para q.

3.7.1. MATRIZ DE PERDIDAS

La definición de las matrices $L_c(q)$ y $L_c(p)$ requiere que las matrices $I_o(q)$ e $I_o(p)$ sean no singulares. Los corolarios 3.6.10 y 3.6.15 garantizan este hecho.

Comenzaremos calculando $L_c(q)$, debido a su sencillez; posteriormente obtendremos $L_c(p)$ a partir de $L_c(q)$.

Resultado 3.7.1

La matriz $L_c(q)$ es diagonal para toda c y $q \in Q$, y sus elementos diagonales

$$L_c^{ii}(q) = 1-c_{(i)} \qquad i=1,\ldots,k,$$

son las pérdidas relativas de información acerca de cada q_i, $i=1,\ldots,k$.

Demostración

Por definición

$$L_c(q) = I - I_c(q)I_o^{-1}(q) \qquad \text{(Definición 2.3.1)}$$

Por ser $I_c(q)$ diagonal para toda c (resultado 3.4.6), la matriz $L_c(q)$ es diagonal con elemento i-ésimo la pérdida relativa de información acerca de q_i (corolario 2, sección 2.3).

Teniendo en cuenta el corolario 3.4.6, obtenemos el resultado.

□

Llamamos la atención sobre el hecho interesante de que la matriz de pérdidas $L_c(q)$ no depende del parámetro q, únicamente depende de la distribución de la censura c.

Por otro lado, todas las medidas reales de la pérdida relativa de información, basadas en la matriz $L_c(q)$ son invariantes para transformaciones biyectivas del parámetro q; ésto hace que resulte innecesaria la obtención de la matriz $L_c(p)$. De esta forma, el siguiente resultado, que recoge la forma de dicha matriz $L_c(p)$ carece de utilidad posterior y se incluye sólo a modo de complemento de la matriz $L_c(q)$ y con una finalidad ilustrativa exclusivamente.

Resultado 3.7.2

La matriz $L_c(p)$ es triangular superior para toda c y $p \in P$ siendo sus elementos

$$L_c^{ii}(p) = 1-c_{(i)} \qquad i=1,\dots,k$$

$$L_c^{ij}(p) = -\frac{(1-c_{(i)})p_j}{p_{(i+1)}} - \sum_{\ell=i+1}^{j-1} \frac{(1-c_{(\ell)})p_\ell p_j}{p_{(\ell)}p_{(\ell+1)}} + \frac{(1-c_{(j)})p_j}{p_{(j)}} \qquad j>i$$

Demostración

$$L_c(p) = A' L_c(q) A'^{-1} \qquad \text{(teorema 3, sección 2.3)}$$

siendo A la matriz triangular inferior definida en el resultado 3.5.1.

$A'L_c(q)$ será una matriz triangular superior por ser el producto de una matriz triangular superior y otra matriz diagonal. Además A'^{-1} es triangular superior por serlo A'. De esta forma

$L_c(p)$ es una matriz triangular superior por ser el producto de dos matrices triangulares superiores $A'L_c(q)$ y A'^{-1}.

Es sencillo comprobar que

$$A'^{-1} = \begin{pmatrix} 1 & -\frac{p_2}{p_{(2)}} & -\frac{p_3}{p_{(2)}} & -\frac{p_4}{p_{(2)}} & \cdots\cdots\cdots & -\frac{p_k}{p_{(2)}} \\ 0 & p_{(2)} & -\frac{p_3 p_{(2)}}{p_{(3)}} & -\frac{p_4 p_{(2)}}{p_{(3)}} & \cdots\cdots\cdots & -\frac{p_k p_{(2)}}{p_{(3)}} \\ 0 & 0 & p_{(3)} & -\frac{p_4 p_{(3)}}{p_{(4)}} & \cdots\cdots\cdots & -\frac{p_k p_{(3)}}{p_{(4)}} \\ 0 & 0 & 0 & p_{(4)} & \cdots\cdots\cdots & -\frac{p_k p_{(4)}}{p_{(5)}} \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & 0 & \cdots\cdots\cdots & p_{(k)} \end{pmatrix}$$

El elemento (i,j) de $L_c(p)$ será el producto de la fila i-ésima de la matriz $A'L_c(q)$ por la columna j-ésima de A'^{-1}.

$\square$

Los siguientes resultados proporcionan las expresiones de las medidas reales $L_c^i(q)$ $i=1,2,3$, de la pérdida relativa de información, acerca de q, debida a la censura. El cálculo de sus homólogos $L_c^i(p)$ $i=1,2,3$, acerca de p, es innecesario ya que, como indicábamos antes, todas estas medidas son invariantes para transformaciones biyectivas del parámetro.

Resaltamos el hecho de que dichas medidas no dependen del parámetro q, dependiendo únicamente de la distribución de la censura c, como consecuencia del mismo hecho para la matriz $L_c(q)$.

Resultado 3.7.3

$$L_c^1(q) = \frac{\sum_{i=1}^{k} (1-c_{(i)})}{k} \qquad \text{para todo } q \in Q$$

Demostración

Por definición

$$L_c^1(q) = \frac{1}{k}\,\text{Traza}\left[L_c(q)\right] \qquad \text{(definición 2.3.2)}$$

el resultado 3.7.1 completa la demostración.

□

$L_c^1(q)$ es, por definición, la media aritmética de las pérdidas relativas de información acerca de cada q_i, $i=1,\dots,k$.

Resultado 3.7.4

$$L_c^2(q) = \prod_{i=1}^{k} (2-c_{(i)})^{1/k} - 1 \qquad \text{para todo } q \in Q$$

Demostración

Por definición

$$L_c^2(q) = \left[\prod_{i=1}^{k} \left[1+\lambda_i(L_c(q))\right]\right]^{1/k} - 1 \qquad \text{(definición 2.3.3)}$$

Por ser $L_c(q)$ diagonal, sus autovalores son sus elementos diagonales. El resultado 3.7.1 completa la demostración.

□

$L_c^2(q)$ es, por definición, la φ-media de las pérdidas relativas de información acerca de cada q_i, $i=1,\dots,k$, siendo $\varphi(x) = \log(1+x)$.

Resultado 3.7.5

$$L_c^3(q) = \left[\frac{\sum_{i=1}^{k} (1-c_{(i)})^2}{k}\right]^{1/2} \qquad \text{para todo } q \in Q$$

Demostración

Por definición

$$L_c^3(q) = \left[\frac{\sum_{i=1}^{k}\left[\lambda_i(L_c(q))\right]^2}{k}\right]^{1/2} \qquad \text{(definición 2.3.4)}$$

Basta sustituir el resultado 3.7.1.

□

$L_c^3(q)$ es, por definición, la media cuadrática de las pérdidas relativas de información acerca de cada q_i, $i=1,\ldots,k$.

Como resumen de lo expuesto en el apartado 3.7.1 destacamos tres hechos:

1) Tanto la matriz de pérdidas $L_c(q)$ como sus medidas reales asociadas $L_c^i(q)$ $i=1,2,3$ dependen únicamente de la distribución c de la variable de censura, y no del parámetro q.

2) Las medidas reales $L_c^i(q)$ $i=1,2,3$ son tres promedios diferentes de las pérdidas relativas de información que, acerca de cada q_i, $i=1,\ldots,k$, origina la censura. De esta forma, la discusión en cuanto a la elección de una u otra de estas tres medidas se reduce, en cada caso, al análisis de la representatividad, ventajas e inconvenientes de los promedios correspondientes.

3) Dichas medidas se encuentran relacionadas (ver observación 2, sección 2.3) por las siguientes desigualdades

$$L_c^2(q) \leq L_c^1(q) \leq L_c^3(q) \quad \text{para todo } q \in Q$$

dándose las dos igualdades a la vez si y sólo si, para c fija, la pérdida relativa de información acerca de q_i es constante para todo $i=1,\ldots,k$, lo que equivale (resultado 3.7.1) a que $c_{(i)}=1 \quad i=1,\ldots,k$ en cuyo caso no se censuran las observaciones.

3.7.2. MATRIZ DE EFICIENCIAS

Al igual que ocurría con la matriz de pérdidas, ahora basta obtener la matriz de eficiencias $R_c(q)$ para los fines que perseguimos.

Resultado 3.7.6

La matriz $R_c(q)$ es diagonal para toda c y $q \in \mathcal{Q}$ y sus elementos diagonales

$$R_q^{ii}(c) = c_{(i)} \qquad i=1,\ldots,k$$

son las eficiencias relativas, acerca de cada q_i, $i=1,\ldots,k$.

Demostración

Por definición

$$R_c(q) = I-L_c(q) \qquad \text{(definición 2.4.1)}$$

La expresión numérica se obtiene inmediatamente del resultado 3.7.1; el corolario 6.2 (sección 2.4) completa la demostración.

$\square$

La matriz $R_c(p)$ se obtiene de forma análoga a partir de $L_c(p)$, o bien a partir de $R_c(q)$ mediante la expresión $R_c(p) = A'R_c(q)A'^{-1}$.

Los siguientes tres resultados muestran los valores de las medidas reales $R_c^i(q)$ $i=1,2,3$ de la eficiencia relativa de ε_c comparado con ε_o, acerca de q.

<u>Resultado 3.7.7</u>

$$R_c^1(q) = \frac{\sum_{i=1}^{k} c_{(i)}}{k} \qquad \text{para todo } q \in Q$$

<u>Demostración</u>

Por definición

$$R_c^1(q) = \frac{1}{k} \text{ Traza}\left[R_c(q)\right] \qquad \text{(definición 2.4.2)}$$

el resultado 3.7.6 completa la demostración.

□

$R_c^1(q)$ es, por definición, la media aritmética de las eficiencias relativas, acerca de cada q_i, $i=1,\ldots,k$.

<u>Resultado 3.7.8</u>

$$R_c^2(q) = \prod_{i=1}^{k} (1+c_{(i)})^{1/k} -1 \qquad \text{para todo } q \in Q$$

<u>Demostración</u>

Por definición

$$R_c^2(q) = \left[\prod_{i=1}^{k} \left[1+\lambda_i(R_c(q))\right]\right]^{1/k} -1 \qquad \text{(definición 2.4.3)}$$

el resultado 3.7.6 completa la demostración.

□

$R_c^2(q)$ es, por definición, la φ-media de las eficiencias relativas acerca de cada q_i, $i=1,\ldots,k$, siendo $\varphi(x) = \log(1+x)$.

Resultado 3.7.9

$$R_c^3(q) = \left[\frac{\sum_{i=1}^{k} c^2_{(i)}}{k}\right]^{1/2} \qquad \text{para todo } q \in Q$$

Demostración

Por definición

$$R_c^3(q) = \left[\frac{\sum_{i=1}^{k} \left[\lambda_i(R_c(q))\right]^2}{k}\right]^{1/2} \qquad \text{(definición 2.4.4)}$$

Basta sustituir el resultado 3.7.6.

□

$R_c^3(q)$ es, por definición, la media cuadrática de las eficiencias relativas acerca de cada q_i, $i=1,\ldots,k$.

La estrecha relación que existe entre las matrices de pérdidas y de eficiencias hace que las propiedades de ésta sean consecuencia de las de aquélla. Resumimos el apartado 3.7.2 en los siguientes hechos:

1) Tanto la matriz de eficiencias $R_c(q)$ como sus medidas reales asociadas $R_c^i(q)$ $i=1,2,3$ dependen únicamente de la distribución c de la variable de censura, y no del parámetro q.

2) Las medidas reales $R_c^i(q)$ $i=1,2,3$ son tres promedios diferentes de las eficiencias relativas, acerca de q_i, $i=1,\ldots,k$, de ε_c comparado con ε_o.
La elección de la medida real idónea de la eficiencia relativa se reduce, de nuevo, a la elección, en cada caso,

del promedio más representativo.

3) Estas medidas se encuentran relacionadas (ver observación 2, apartado 2.4.3) por las siguientes desigualdades

$$R_c^2(q) \leq R_c^1(q) \leq R_c^3(q) \qquad \text{para todo } q \in Q$$

dándose las dos igualdades a la vez si y sólo si, para c fija, la eficiencia relativa acerca de q_i es constante para todo $i=1,\dots,k$, lo que equivale (resultado 3.7.6) a que $c_{(i)}=1$ $i=1,\dots,k$, en cuyo caso no se censuran las observaciones.

3.8. PERDIDA DE INFORMACION. MEDIDAS REALES

Al comienzo de la sección anterior indicábamos dos métodos para evaluar la pérdida de información debida a la censura. En dicha sección se trató el que denominábamos método matricial. Nos ocuparemos ahora del método real.

Supondremos que las medidas de información acerca del parámetro son las medidas reales S_c, D_c y M_c (basadas en la matriz de Fisher I_c) cuyos cálculos aparecen en la sección 3.6. Para cada una de estas medidas reales definimos de modo natural la pérdida relativa de información y la eficiencia relativa (sección 2.5). Siguiendo la notación del capítulo II, utilizaremos $L_c^S(p)$, $L_c^D(p)$ y $L_c^M(p)$ para denotar la pérdida relativa de información que se produce acerca de p al observar ε_c en vez de ε_o, cuando las medidas de información elegidas son S_c, D_c y M_c respectivamente. Igualmente $R_c^S(p)$, $R_c^D(p)$ y $R_c^M(p)$ denotará la eficiencia relativa acerca de p del experimento ε_c comparado con

ε_o, según las medidas de información respectivas. La letra q en lugar de p expresará las mismas medidas de la pérdida de información anteriores, respecto del parámetro q.

A excepción de la medida L_c^D y como consecuencia R_c^D, las demás medidas no son invariantes para transformaciones biyectivas del parámetro. Esto hace que, en el cálculo de dichas medidas distingamos entre los parámetros p y q. Por su parte, la medida L_c^D es invariante para este tipo de transformaciones bajo hipótesis restrictivas (teorema 8d), sección 2.5) que, en particular, verifica nuestro modelo.

3.8.1. MEDIDA DE INFORMACION S_c

Resultado 3.8.1

$$L_c^S(q) = 1- \sum_{i=1}^{k} c_{(i)} \frac{\dfrac{(1-q_{i-1})\dots(1-q_1)}{q_i(1-q_i)}}{\sum\limits_{i=1}^{k} \dfrac{(1-q_{i-1})\dots(1-q_1)}{q_i(1-q_i)}}$$

$$R_c^S(q) = 1-L_c^S(q) \qquad \text{para toda c.}$$

Demostración

Por definición

$$L_c^S(q) = 1- \frac{S_c(q)}{S_o(q)} \qquad \text{(expresión (2.5.1), epígrafe 2.5.1.}$$

y

$$R_c^S(q) = 1-L_c^S(q) \qquad \text{(expresión (2.5.4), epígrafe 2.5.1.}$$

El resultado 3.6.3 y el corolario 3.6.4 completan la demostración.

□

Observación

Si tenemos en cuenta el corolario 3.4.7, las medidas $L_c^S(q)$ y $R_c^S(q)$ pueden expresarse en la forma siguiente:

$$L_c^S(q) = \sum_{i=1}^{k} (1-c_{(i)}) \frac{I_o^{ii}(q)}{\sum_{i=1}^{k} I_o^{ii}(q)} \qquad \text{(expresión (2.5.2), epígrafe 2.5.1.1)}$$

$$R_c^S(q) = \sum_{i=1}^{k} c_{(i)} \frac{I_o^{ii}(q)}{\sum_{i=1}^{k} I_o^{ii}(q)} \qquad \text{(expresión (2.5.5), epígrafe 2.5.1.2)}$$

con lo que su interpretación resulta sencilla:

$L_c^S(q)$ es la media aritmética ponderada de las pérdidas relativas de información acerca de cada q_i, $i=1,\dots,k$, siendo las ponderaciones respectivas las proporciones de información de Fisher acerca de cada q_i según el experimento ε_o.

$R_c^S(q)$ es la media aritmética ponderada de las eficiencias relativas acerca de cada q_i, $i=1,\dots,k$, con las mismas ponderaciones que antes.

Resultado 3.8.2

$$L_c^S(p) = \sum_{i=1}^{k} c_i - \sum_{i=1}^{k} c_i \frac{\dfrac{i}{p_{(i+1)}} + \sum_{j=1}^{i} \dfrac{1}{p_j}}{\dfrac{k}{p_{k+1}} + \sum_{i=1}^{k} \dfrac{1}{p_i}}$$

$$R_c^S(p) = 1-L_c^S(p) \qquad \text{para toda } c.$$

Demostración

Análoga a la anterior. Ahora el resultado 3.6.1 y el corolario 3.6.2 completan la demostración.

□

3.8.2. MEDIDA DE INFORMACION D_c

La definición de la medida $L_c^D(q)$ exige que la matriz $I_o(q)$ sea no singular (epígrafe 2.5.2.1). El corolario 3.6.10 garantiza este hecho.

Además, para c prefijada, esta medida $L_c^D(q)$ es invariante para transformaciones biyectivas f(q) del parámetro q, si la matriz de derivadas parciales $(\frac{\partial q_i}{\partial f_j})$ i,j=1,...,k es no singular (teorema 8 d), sección 2.5). Por ser $A=(\frac{\partial q_i}{\partial p_j})$ no singular (observación del corolario 3.5.3) la propiedad de invariancia hace que $L_c^D(q) = L_c^D(p)$. Calculamos esta última por su simplicidad.

<u>Resultado 3.8.3</u>

$$L_c^D(p) = 1-(\prod_{i=1}^{k} c_{(i)})^{\alpha} \quad \text{para todo } p \in P$$

$$R_c^D(p) = (\prod_{i=1}^{k} c_{(i)})^{1/k} \quad \text{para todo } p \in P$$

<u>Demostración</u>

Por definición

$$L_c^D(p) = 1- \frac{D_c(p)}{D_o(p)} \quad \text{(expresión (2.5.6), epígrafe 2.5.2.1}$$

Además

$$R_c^D(p) = \left[1-L_c^D(p)\right]^{1/k\alpha} \quad \text{(expresión (2.5.8), epígrafe 2.5.2.2}$$

El resultado 3.6.11 y el corolario 3.6.12 completan la demostración.

□

Observación (ver observación 1, apartado 2.5.2)

Para $\alpha = \frac{1}{k}$ observamos que $L_c^D(p)$ es la φ-media de las pérdidas relativas de información acerca de cada q_i, $i=1,\dots,k$, siendo $\varphi(x)=\log(1-x)$ (Calot; 1970, p. 73).

Para todo $\alpha>0$ que fijemos, $R_c^D(p)$ es la media geométrica de las eficiencias relativas acerca de cada q_i, $i=1,\dots,k$.

3.8.3. MEDIDA DE INFORMACION M_c

Resultado 3.8.4

$$L_c^M(q) = 1 - \left[\sum_{i=1}^{k} c_{(i)}^2 \frac{\dfrac{(1-q_{i-1})^2\dots(1-q_1)^2}{q_i^2(1-q_i)^2}}{\displaystyle\sum_{i=1}^{k} \frac{(1-q_{i-1})^2\dots(1-q_1)^2}{q_i^2(1-q_i)^2}} \right]^{1/2}$$

$R_c^M(q) = 1-L_c^M(q)$ para toda c.

Demostración

Por definición

$$L_c^M(q) = 1- \frac{M_c(q)}{M_o(q)}$$ (expresión (2.5.9), epígrafe 2.5.3.1)

y

$$R_c^M(q) = 1 - L_c^M(q)$$ (expresión (2.5.11) , epígrafe 2.5.3.2

El resultado 3.6.19 y el corolario 3.6.20 completan la demostración.

□

<u>Observación</u> (ver observación 1, apartado 2.5.3)

Teniendo en cuenta el corolario 3.4.7, las medidas $L_c^M(q)$ y $R_c^M(q)$ pueden expresarse en la forma siguiente:

$$L_c^M(q) = 1 - \left[\sum_{i=1}^{k} c_{(i)}^2 \frac{(I_o^{ii}(q))^2}{\sum_{i=1}^{k} (I_o^{ii}(q))^2} \right]^{1/2}$$

$$R_c^M(q) = \left[\sum_{i=1}^{k} c_{(i)}^2 \frac{(I_o^{ii}(q))^2}{\sum_{i=1}^{k} (I_o^{ii}(q))^2} \right]^{1/2}$$

con lo que su interpretación resulta sencilla:

$L_c^M(q)$ es la φ-media ponderada de las pérdidas relativas de información acerca de cada q_i, $i=1,\dots,k$, siendo las ponderaciones respectivas las proporciones de las informaciones de Fisher al cuadrado, acerca de cada q_i según el experimento ε_o; y $\varphi(x)=(1-x)^2$ (Calot; 1970, p. 73).

$R_c^M(q)$ es la media cuadrática ponderada de las eficiencias relativas acerca de cada q_i, $i=1,\dots,k$, con las mismas ponderaciones que antes.

El cálculo de las medidas $L_c^M(p)$ y $R_c^M(p)$ resulta sencillo a partir del resultado 3.6.17 y del corolario 3.6.18. Omitimos, sin embargo, dichos cálculos debido a que la complejidad de las expresiones obtenidas resulta escasamente ilustrativa.

Como resumen de lo expuesto en la sección 3.8 destacamos lo siguiente:

1) Las medidas de la pérdida relativa de información $L_c^S(q)$ y $L_c^M(q)$ son dos promedios ponderados diferentes de las pérdidas relativas de información que, acerca de cada q_i, $i=1,\dots,k$, origina la censura.
En el caso de la medida $L_c^S(q)$ las ponderaciones son directamente proporcionales a las informaciones de Fisher, acerca de cada q_i, según el experimento ε_o. De esta forma, las pérdidas relativas de información que tendrán mayor influencia en la medida L_c^S serán las relativas a componentes q_i para las que el experimento ε_o aporte mayor información. La medida $L_c^M(q)$ potencia aún más la influencia de las pérdidas relativas de información anteriores, al ser las ponderaciones correspondientes proporcionales a los cuadrados de las informaciones de Fisher acerca de cada q_i, según ε_o.
En base a esta explicación, parecen más precisas como medidas de la pérdida relativa de información, acerca de q, las medidas $L_c^S(q)$ y $L_c^M(q)$ que $L_c^1(q)$ y $L_c^3(q)$ obtenidas por el método matricial, sobre todo en aquellas situaciones en que las informaciones de Fisher acerca de las q_i sean muy dispares. Esta ventaja intuitiva de las medidas "reales" respecto de las "matriciales" tiene su contrapartida formal en la propiedad de invariancia para transformaciones biyectivas del espacio paramétrico, que verifican las segundas y no las primeras.

2) Las pérdidas relativas de información, acerca de cada q_i, $1-c_{(i)}$, sólo dependen de la distribución c de la variable de censura. Si promediamos dichas pérdidas mediante funcio

nes del parámetro q, los promedios correspondientes, en este caso $L_c^S(q)$ y $L_c^M(q)$, son funciones de c y q.

3) La medida $L_c^S(p)$ es, por definición, la media aritmética ponderada de las pérdidas relativas de información que, acerca de cada p_i, $i=1,\ldots,k$, origina la censura, siendo las ponderaciones proporcionales a las informaciones de Fisher, acerca de cada p_i, según el experimento ε_o.
La medida $L_c^M(p)$ no puede expresarse formalmente como un promedio ponderado de las pérdidas relativas de información, acerca de cada p_i, $i=1,\ldots,k$.

4) Como veíamos en el capítulo II (observación 3, apartado 2.5.2) es necesario exigir la no singularidad de la matriz de Fisher I_c para todo valor del parámetro, para garantizar el buen comportamiento de la medida L_c^D. En nuestro caso la no singularidad de las matrices $I_c(p)$ e $I_c(q)$ equivale a que $c_{(k)}\neq 0$ (corolarios 3.6.9 y 3.6.14). Sólo cuando la variable de censura verifique dicha condición debemos utilizar L_c^D como medida de la pérdida relativa de información. En este caso, y tomando $\alpha = \frac{1}{k}$, la medida $L_c^D(q)$ es un promedio (no ponderado) de las pérdidas relativas de información acerca de cada q_i, $i=1,\ldots,k$, por lo que sólo depende de la distribución c de la variable de censura y no del parámetro q.
Además se verifica que

$L_c^D(q) \geqslant L_c^2(q)$ para toda c y $q \in Q$ (observación 2, apartado 2.5.2)

<u>con la igualdad si y sólo si</u>, para c fija, la pérdida relativa de información acerca de q_i es constante para todo $i=1,\ldots,k$, lo que equivale (resultado 3.7.1) a que $c_{(i)}=1$, $i=1,\ldots,k$, en cuyo caso <u>no se censuran las observaciones</u>.

5) Un resumen similar, como el expuesto en los cuatro puntos anteriores, pero referido a la eficiencia relativa, podría hacerse para las medidas R_c^S, R_c^D y R_c^M .

APENDICE

APENDICE

Sea A_k el conjunto de matrices reales semidefinidas positivas de orden k.

Definimos en A_k el orden parcial (Reflexiva, Antisimétrica y Transitiva) usual, es decir: $A,B \in A_k$, $A \geq B \iff A-B \in A_k$

Para cualquier $A \in A_k$, sean $\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_k(A) \geq 0$ los autovalores A en orden de magnitud decreciente.

A.1. Lema (Okamoto y Kanazawa; 1968).

Una condición necesaria y suficiente para que una función real f(A) definida sobre A_k sea:

i) estrictamente creciente, es decir $f(A) \geq f(B)$ si $A \geq B$ y $f(A) > f(B)$ si además $A \neq B$, y

ii) invariante bajo transformaciones ortogonales, es decir $f(P'AP)=f(A)$ para cualquier matriz ortogonal P

es que f(A) sea idéntica a alguna función $g(\lambda_1(A),\ldots,\lambda_k(A))$ de los autovalores de A estrictamente creciente en cada argumento.

A.2. Propiedad

Para toda $A \in A_k$ se cumple

A es definida positiva $\iff$ A es no singular

Demostración

$A \in A_k \Rightarrow \lambda_i(A) \geq 0 \quad i=1,\ldots,k$

Además $|A| = \prod_{i=1}^{k} \lambda_i(A) \geq 0$ entonces

$|A| \neq 0 \iff \lambda_i(A) > 0$, $i=1,\ldots,k$ $\iff$ A es definida positiva.

$\square$

REFERENCIAS BIBLIOGRAFICAS

REFERENCIAS BIBLIOGRAFICAS

Anderson, T.W. y Das Gupta, S. (1963). Some inequalities on characteristic roots of matrices. Biometrika 50, 522-524.

Apostol, T.M. (1986). Análisis Matemático. 2ª edición. Reverté, s.a.

Barlow, R. y Hsiung, J. (1983). Information in a life test experiment. The Statistician 32, 35-45.

Bayarri, M.J. y DeGroot, M.H. (1986). Information in selection models. Department of Statistics. Carnegie-Mellon University. Technical Report 368.

Bellman, R. (1970). Introduction to matrix analysis. McGraw-Hill. New York.

Bernardo, J.M. (1979). Expected information as expected utility. Ann. Statist, 7, 686-690.

Blackwell, D. (1951). Comparison of experiments. Proceedings of the Second Berkeley Symposium on Mathematical Stat. and Prob. University of California Press, 93-102.

Blackwell, D. (1953). Equivalent comparisons of experiments. Ann. Math. Statist. 24, 265-272.

Boekee, D.E. (1977). An extension of the Fisher information measure in "Topics in Information Theory" (I. Csiszár and P. Elias, Eds.), 113-123, North-Holland, Amsterdam.

Brooks, R.J. (1980). On the relative efficiency of two paired--data experiments. J. Roy. Statist. Soc. B. 42, 186--191.

Brooks, R.J. (1982). On the loss of information through censoring. Biometrika, 69, 137-144.

Calot, G. (1970). Curso de estadística descriptiva. Paraninfo, Madrid.

Cox, D.R. y Oakes, D. (1984). Analysis of survival data. Chapman and Hall, London.

Csiszár, I. (1963). Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von Markoffschen Ketten, Mayar Tud. Akad Mat. Kutató Int. Kozl.

DeGroot, M.H. (1962). Uncertainty, information and sequential experiments. Ann. Math. Statist. 33, 404-419.

DeGroot, M.H. (1966). Optimal allocation of observations. Annals Inst. Statist. Math. 18. 13-28.

DeGroot, M.H. (1984). Changes in utility as information. Theory and Decision 17, 283-303.

Faddeev, D.K. y Faddeeva, V.N. (1963). Computational methods of linear algebra (Edición Revolucionaria). Instituto cubano del libro.

Ferentinos, K. y Papaioannou, T. (1979). Loss of information due to groupings. Trans. of the Eighth Prague Conference on Information Theory. Statistical Decision Functions and Random Processes, 1978, C, 87-94.

Ferentinos, K. y Papaioannou, T. (1981). New parametric measures of information. Information and Control, 51, 193--208.

Ferentinos, K. y Papaioannou, T. (1982). Information in experiments and sufficiency. Jour. Statist. Planning Inference, 6, 309-317.

Ferentinos, K. y Papaioannou, T. (1983). Convexity of measures of information and loss of information due to grouping of observations. Jr. Comb. Inf. & Sci. Vol. 8, n° 4, 286-294.

Fisher, R.A. (1925). Theory of statistical estimation. Proc. Cambridge. Philos. Soc. 22, 700-725.

Fisher, R.A. (1956). Statistical methods and scientific inference. Oliver and Boyd. Edinburgh.

Fourgeaud, C. y Fuchs, A. (1972). Statistique. Dunod.

Goel, P. (1987). Comparison of experiments and information in censored data. Proceedings of 4th Purdue Symposium on Statistical Decision Theory and Relation Topics, Vol. 2, Academic Press. To appear.

Goel, P. y DeGroot, M. (1979). Comparison of experiments and information measures. Ann. Statist. 7, 1066-1077.

Hansen, O.H. y Torgersen, E.N. (1974). Comparison of linear normal experiments. Ann. Statist. 2, 367-373.

Hodges, J.L., Jr. y Lehmann, E.L. (1956). The efficiency of some nonparametric competitors of the t-test. Ann. Math. Statist. 27, 324-335.

Hollander, M., Proschan, F. y Sconing, J. (1985 a). Information in censored models. Tech. Report. M 701, Department of Statistics, Tallahassee, Florida State University.

Hollander, M., Proschan, F. y Sconing, J. (1985 b). Measures of dependence for evaluating information in censored models. Tech. Report. M 706, Department of Statistics, Tallahassee, Florida State University.

Kagan, A.M. (1963). On the theory of Fisher's amount of information. Sov. Math. Dokl. 4, 991-993.

Kagan, A.M., Linnik, Y.V. y Rao, C.R. (1973). Characterization problems in mathematical statistics, Wiley, New York.

Kalbfleich, J.D. y Prentice, R.L. (1980). The statistical analysis of failure time data. New York: Wiley.

Kullback, S. (1968). Information theory and statistics. Dover. New York.

Kullback, S. y Leibler, A. (1951). On information and sufficiency. Ann. Math. Statist. 22, 79-86.

Lee, Elisa T. (1980). Statistical methods for survival data analysis. L.L.P. Belmont. California.

Lehmann, E.L. (1986). Testing statistical hypotheses. 2nd Ed. New York, Wiley.

Lindley, D.V. (1956). On a measure of the information provided by an experiment. Ann. Math. Statist. 27, 986-1005.

Mallows, C. (1959). The information in an experiment. J. Roy. Statist. Soc. Ser. B. 21, 67-72.

Mardia, K.V., Kent, J.T. y Bibby, J.M. (1979). Multivariate analysis. Academic Press, London.

Mathai, A. (1967). Dispersion and information. Metron 26, 1-12.

Matusita, K. (1967). On the notion of affinity of several distributions and some of its applications. Ann. Inst. Statist. Math. 19, 181-192.

Miller, R.G., Jr. (1981). Survival analysis. New York, Wiley.

Nelson, W. (1982). Applied life data analysis. New York, Wiley.

Okamoto, M. y Kanazawa, M. (1968). Minimization of eigenvalues of a matrix and optimality of principal components. Ann. Math. Statist. 39, 859-863.

Papaioannou, T. y Kempthorne, O. (1971). On statistical information theory and related measures of information. Aerospace Research Laboratories Report, ARL 71-0059, Wright--Patterson A.F.B., Ohio.

Raiffa, H. y Schlaifer, R. (1961). Applied statistical decision theory. The M.I.T. Press.

Rao, C.R. (1973). Linear statistical inference and its applications. Wiley, New York.

Rényi, A. (1961). On measures of entropy and information. Berkeley Symposium on Math. Statist. and Prob. Vol. I, 547--561. Univ. Of California Press, Berkeley.

Rohatgi, V.K. (1976). An introduction to probability theory and mathematical statistics. New York, Wiley.

Shannon, C.E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423, 623-656.

Stam, A. (1959). Some mathematical properties of quantities of information. Ph. D. thesis. Delft Univ. of Technology, Delft, the Netherlands.

Stone, M. (1961). Non-equivalent comparison of experiments and their use for experiments involving location parameters. Ann. Math. Statist. 32, 326-332.

Vajda, I. (1973). χ^2-divergence and generalized Fisher's information. Transactions of the 6th Prague Conference on Information Theory, Stat. Dec. Func. and Random Proc., 1971, 873-886.

Wilks, S.S. (1962). Mathematical statistics. Wiley, New York.